

GenCore version 5.1.3  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - protein search, using frame\_plus\_n2p model

Run on: February 12, 2003, 08:54:59 ; Search time 14 Seconds

(without alignments)  
3750.645 Million cell updates/sec

Title: US-09-899-303a-7

Perfect score: 1155  
Sequence: 1 ATGTTGGTGAAGTCATCGA.....TGATGATGAACGTGTAATAG 633

Scoring table: BIOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 112892 segs, 41476328 residues

Total number of hits satisfying chosen parameters: 225784

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Command line parameters:  
-MODE=frame+n2p.model -DEV=x1p  
-O=/cgn2.1/USPTO.spool/US09899303/rnat.12022003.085448.28260/app.query.fasta.1.775  
-DB=SwissProt.40 -OPMT=fastan -SUFFIX=n2p.rsp -MINMATCH=0.1 -LOOPCL=0  
-LOOPEXT=0 -UNIT=bits -START=1 -END=1 -MATRIX=blomsun62 -TRANS=human40.cdl  
-LIST=45 -DOCALLIN=200 -THR\_SCORE=pct -THR\_MAX=100 -THR\_MIN=0 -ALIGN=15  
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEADSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US09899303.ecgn.1.19.rnat.12022003.085448.28260 -NCPU=6 -ICPU=3  
-NO\_XLPHY -NO\_MMAP -LARGEQUERY -NEG\_SCORES=0 -WAIT -LONGLOG -DEV\_TIMEBOOT=120  
-WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7  
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database: SwissProt\_40.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1078	93.3	3010	1	POLG_HCVTW
2	1073	92.9	513	1	POLG_HCVJ2
3	1073	92.9	3010	1	POLG_HCVJA
4	1073	92.9	3010	1	POLG_HCVJT
5	1060	91.8	3010	1	POLG_HCVBK
6	1057	91.5	520	1	POLG_HCVHK
7	1044	90.4	520	1	POLG_HCVH4
8	995	86.1	3011	1	POLG_HCVH
9	992	85.9	3011	1	POLG_HCV1
10	988	85.5	321	1	POLG_HCVTH
11	983	85.1	321	1	POLG_HCVH8
12	978	80.3	309	1	POLG_HCVH7
13	878	76.0	192	1	POLG_HCVH1
14	826	71.5	737	1	POLG_HCVJ7
15	817	70.7	3033	1	POLG_HCVJ8
16	811	70.2	737	1	POLG_HCVJ5
17	802	69.4	3033	1	POLG_HCVJ6
18	155	13.4	138	1	POLG_HCVB0

C	19	97	8.6	305	1	YT32_STRFR	P20185 streptomyc
C	20	94	8.1	2476	1	ZAN_PIG	Q28983 sus scrofa
C	21	91.5	8.1	483	1	VE2_HPV14	P36783 human papil
C	22	90	8.0	400	1	B3AR_RAT	P26255 rattus norv
C	23	89.5	8.0	408	1	B3AR_HUMAN	P13945 homo sapien
C	24	89.5	7.7	1125	1	MAP4_MOUSE	P27546 mus musculu
C	25	88	7.8	400	1	B3AR_MOUSE	P25962 mus musculu
C	26	87.5	7.8	493	1	VE2_HPV19	P36786 human papil
C	27	87	7.7	2426	1	SON_HUMAN	P18583 homo sapien
C	28	86.5	7.7	1415	1	ICP4_HSVMG	P02362 marek's dis
C	29	85	7.6	405	1	B3AR_SHEEP	O9X558 ovis aries
C	30	84.5	7.3	821	1	CAN3_RAT	P16259 rattus norv
C	31	84	7.5	358	1	B3AR_FELCA	O91614 felis silve
C	32	84	7.5	1077	1	AT10_HUMAN	O9H324 homo sapien
C	33	84	7.3	1239	1	NME3_MOUSE	O01098 mus musculu
C	34	84	7.3	3014	1	CLRI_HUMAN	O9NY96 homo sapien
C	35	83.5	7.2	810	1	CAN3_CHICK	O92177 gallus galli
C	36	83	7.4	405	1	B3AR_CANFA	O02662 canis fami
C	37	82.5	7.3	342	1	RX_RAT	O9J1E7 rattus norv
C	38	82.5	7.1	821	1	CAN3_MOUSE	O64691 mus musculu
C	39	82.5	7.3	827	1	AXNI_RAT	O70239 rattus norv
C	40	82.5	7.3	2404	1	SON_MOUSE	O9GX47 mus musculu
C	41	82	7.1	198	1	VGLX_HSVEL	P25088 equine herp
C	42	82	7.1	383	1	VGLX_HSVK	P32515 equine herp
C	43	82	7.3	418	1	B3AR_MACMU	Q28524 macaca mula
C	44	82	7.3	676	1	ICP0_HSVBJ	P29128 bovine herp
C	45	82	7.1	797	1	VGLX_HSVBE	P28968 equine herp

#### ALIGNMENTS

RESULT 1  
POLG\_HCVTW  
ID POLG\_HCVTW STANDARD: PRT: 3010 AA.  
AC P29846:  
DT 01-APR-1993 (Rel. 25, Last sequence update)  
DT 01-APR-1993 (Rel. 41, Last annotation update)  
DT 15-JUN-2002 (Rel. 41, Last annotation update)  
DE Genome polyprotein [Contains: Capsid protein C (Core protein) (P22);  
DE Envelope glycoprotein E1 (GP32) (GP35); Envelope glycoprotein E2  
DE (GP68) (GP70); Protein P7; Nonstructural protein NS2 (P21)  
DE (EC 3.4.22.-); Protease/helicase NS3 (P70) (hepacivirin)  
DE (EC 3.4.21.98); Nonstructural protein NS4A (P4); Nonstructural protein  
DE NS4B (P27); Nonstructural protein NS5A (P56); Nonstructural protein  
DE NS5B (P66) (P70) (RNA-directed RNA polymerase) (EC 2.7.7.48)].  
OS Hepatitis C virus (isolate Taiwan) (HCV).  
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;  
OC Hepacivirus.  
OX NCBI\_TaxID=31645;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE-92230206; PubMed-1314449;  
RA Chen P.J., Lin M.H., Tai K.F., Liu P.C., Lin C.J., Chen D.S.;  
RT "The Taiwanese hepatitis C virus genome: sequence determination and  
RT mapping the 5' terminus of viral genomic and antigenomic RNA.";  
RL Virology 188:102-113(1992).  
CC -1- FUNCTION: THE SMALL PROTEINS NS2A, NS2B, NS4A AND NS4B ARE  
CC HYDROPHOBIC, SUGGESTING A POSSIBLE MEMBRANE-RELATED FUNCTION.  
CC NS3 AND NS5 MAY PLAY A ROLE IN THE VIRAL RNA REPLICATION.  
CC -1- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral  
CC precursor polyprotein, commonly with Asp or Glu in the P6  
CC position, Cys or Thr in P1 and Ser or Ala in P1'.  
CC -1- CATALYTIC ACTIVITY: N nucleoside triphosphate -> N diphosphate +  
CC (RNA)(N).  
CC -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A  
CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:  
CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF  
CC PROTEIN C AND MRNA.  
CC -1- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY S29.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its

[illegible][illegible]

RA Okamoto H., Kurai K., Okada S.I., Yamamoto K., Lizuka H.,  
 RA Tanaka T., Fukuda S., Tsuda F., Mishiro S.;  
 RT "Full-length sequence of a hepatitis C virus genome having poor  
 RT homology to reported isolates: comparative study of four distinct  
 RT genotypes.";  
 RL Virology 188:331-341(1992).  
 CC -1- FUNCTION: THE SMALL PROTEINS NS2A, NS2B, NS4A AND NS4B ARE  
 CC HYDROPHOBIC, SUGGESTING A POSSIBLE MEMBRANE-RELATED FUNCTION.  
 CC NS3 AND NS5 MAY PLAY A ROLE IN THE VIRAL RNA REPLICATION.  
 CC -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A  
 CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:  
 CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF  
 CC PROTEIN C AND MRNA.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: D10074; BAA00968.1; -;  
 DR InterPro: IPR002531; HCV\_NSL.  
 DR InterPro: IPR002522; HCV\_Capsid.  
 DR InterPro: IPR002521; HCV\_core.  
 DR InterPro: IPR002519; HCV\_env.  
 DR Pfam: PF01539; HCV\_env; 1.  
 DR Pfam: PF01542; HCV\_core; 1.  
 DR Pfam: PF01543; HCV\_capsid; 1.  
 DR Pfam: PF01560; HCV\_NSL; 1.  
 DR ProDom: PD186062; HCV\_NSL; 1.  
 DR PolyProtein; Glycoprotein; Coat protein; Envelope protein;  
 KW Transmembrane; Nonstructural protein.  
 FT INIT\_MET 1  
 FT CHAIN 1 115  
 FT CHAIN 116 191  
 FT CHAIN 192 383  
 FT CHAIN 384 >513  
 FT TRANSMEM 347 369  
 FT CARBOHYD 196 196  
 FT CARBOHYD 209 209  
 FT CARBOHYD 233 233  
 FT CARBOHYD 234 234  
 FT CARBOHYD 250 250  
 FT CARBOHYD 305 305  
 FT CARBOHYD 417 417  
 FT CARBOHYD 423 423  
 FT CARBOHYD 430 430  
 FT CARBOHYD 448 448  
 FT NON\_TER 513  
 FT SEQUENCE 513 AA; 55704 MW; 943F31E3514CDE3 CRC64;  
 Alignment Scores:  
 Pted. No.: 4,47e-89 Length: 513  
 Score: 1073.00 Matches: 197  
 Percent Similarity: 97.60% Conservative: 6  
 Best Local Similarity: 94.71% Mismatches: 5  
 Query Match: 92.90% Indels: 0  
 DB: 1 Gaps: 0  
 US-09-899-303a-7 (1-633) x POLG\_HCVJ2 (1-513)  
 QY 4 TTGGGTAAAGTCATCGATCCCTTACGTGGGCTTGGCCGACCTCATGGGTACATTCG 63  
 DB 119 LeuGlyValValIleAspThrLeuThrCysGlyPheAlaAspLeuMetGlyTyrIlePro 138  
 QY 64 CTCGTGCGCCGCCCTAGGCGGTGTCTCCAGAGACCTGGCGCATGCGCTCCGGGTTCTG 123  
 DB 139 LeuValIcIylalProIeuGlyGlyAlaAlaArgIalLeuAlaHISglYValArgValIeu 158  
 QY 124 GAAGACGGCGGTGAACATGTGCAACAGGAAATTGGCTGTGCTTTCTATCTTCCTC 183

DB 159 GluAspSerValAsnTyrAlaThrGlyAlaSerProGlyCysSerPheSerIlePheLeu 178  
 QY 184 TTGGCTTACATGTCCTGTCTGACCATTCGACCTCCGCTTATGATGGCGCAACGTGCC 243  
 DB 179 LeuAlaIeuLeuSerCysIleThrIleProAlaSerIalArgIuValArgAsnValSer 198  
 QY 244 GGGATGACCATGTCTACGACGAGCTGCTCCAACTCAAGCATTTGTATGAGCGACGGAC 303  
 DB 199 GlyAlaIlyrHisValThrAsnAspCysSerAsnSerSerIleValIyGlyAlaIaAsp 218  
 QY 304 ATGATCATGACACACCCCGGGTGGCTGCTGCTGGCTGGGAGAACAACTCTCCCGTGC 363  
 DB 219 LeuIleMetHisAlaPheProGlyCysValProCysValArgIuAsnAsnSerSerArgCys 238  
 QY 364 TTGGGTAGCGGTACCCCGACCGCTGACGATAGAACCGACGCGCCACATGACGACATA 423  
 DB 239 TrpValAlaIeuThrProThrLeuAlaIaArgAsnThrSerIleProThrThrIle 258  
 QY 424 CGAGCCGACGTGATTTGCTGCTGGGCGGCTGCTTCTGCTCCGCTATGATGCTGGG 483  
 DB 259 ArgAlrHisValAlaSerLeuIeuValIGlyAlaIaThrPheCysSerAlaMetYrValIGly 278  
 QY 484 GATCTCTGCGGATGTGCTTCTGCTGCTCCAGCTGTTCACACATCTGCGCTCGGCAT 543  
 DB 279 AspIeuCysGlySerValPheLeuValSerGlnIeuPheThrPheSerProArgArgHis 298  
 QY 544 GAGACGGTGGCAGCAGCTCAATTGCTCAATGTATCCGGCCACATACAGGCGACCGTATG 603  
 DB 299 GluThrIeuGlnAspCysAsnCysSerIleThrProGlyHisIeuSerGlyHisArgMet 318  
 QY 604 GCTTGGATATGATGATGAACCTGG 627  
 DB 319 AlaTrpAspMetMetMetAsnTTP 326  
 RESULT 3  
 POLG\_HCVJ2  
 ID POLG\_HCVJ2 STANDARD; PRT; 3010 AA.  
 AC P26662;  
 DT 01-AUG-1992 (Rel. 23, Created)  
 DT 01-AUG-1992 (Rel. 23, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Genome polypeptide [contains: capsid protein C (core protein) (P22);  
 DE Envelope glycoprotein E1 (GP32) (GP35); Envelope glycoprotein E2  
 DE (GP68) (GP70) (NS1); Protein P7; Nonstructural protein NS2 (P21)  
 DE (BC 3.4.22.-); Protease/helase NS3 (P70) (hepativirin)  
 DE (BC 3.4.21.98); Nonstructural protein NS4A (P4); Nonstructural protein  
 DE NS4B (P27); Nonstructural protein NS5A (P56); Nonstructural protein  
 DE NS5B (P66) (P70) (RNA-directed RNA polymerase) (BC 2.7.7.48)].  
 OS Hepatitis C virus (isolate Japanese) (HCV).  
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;  
 OC Hepacivirus.  
 OX NCBI\_TaxID=11116;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=91088550; PubMed=2175903;  
 RA Kato N., Hijikata M., Ootsuyama Y., Nakagawa M., Ohkoshi S.,  
 RA Sugimura T., Shimotohno K.;  
 RT "Molecular cloning of the human hepatitis C virus genome from  
 RT Japanese patients with non-A, non-B hepatitis.";  
 RT Proc. Natl. Acad. Sci. U.S.A. 87:9524-9528(1990).  
 RN [2]  
 RP DISCUSSION OF SEQUENCE.  
 RX MEDLINE=91192160; PubMed=1849488;  
 RA Kato N., Hijikata M., Nakagawa M., Ootsuyama Y., Muraiso K.,  
 RA Ohkoshi S., Shimotohno K.;  
 RT "Molecular structure of the Japanese hepatitis C viral genome.";  
 RT FEBS Lett. 280:325-328(1991).  
 CC -1- FUNCTION: THE SMALL PROTEINS NS2A, NS2B, NS4A AND NS4B ARE  
 CC HYDROPHOBIC, SUGGESTING A POSSIBLE MEMBRANE-RELATED FUNCTION.  
 CC NS3 AND NS5 MAY PLAY A ROLE IN THE VIRAL RNA REPLICATION.  
 CC -1- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral  
 CC precursor polypeptide, commonly with Asp or Glu in the P6



DT 15-JUN-2002 (Rel. 41, last annotation update)  
 DE Genome polypeptide [Contains: Capsid protein C (core protein) (P22);  
 DE Envelope glycoprotein E1 (GP32) (GP35); Envelope glycoprotein E2  
 DE (GP68) (GP70) (NS1); Protein P7; Nonstructural protein NS2 (P21)  
 DE (BC 3.4.22.-); Protease/helicase NS3 (P70) (Hepacivirin)  
 DE (BC 3.4.21.98); Nonstructural protein NS4A (P4); Nonstructural protein  
 DE NS4B (P27); Nonstructural protein NS5A (P56); Nonstructural protein  
 DE NS5B (P66) (P70) (RNA-directed RNA polymerase) (BC 2.7.7.48)].  
 OS Hepatitis C virus (isolate HC-J7) (HCV).  
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;  
 OC Hepacivirus.  
 OX NCBI\_TaxID:31642;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE:92295714; PubMed:1318627;  
 RA Tanaka T., Kato N., Nakagawa M., Ootsuyama Y., Cho M.J.,  
 RA Nakazawa T., Hijioka M., Ishimura Y., Shimotohno K.;  
 RT Molecular cloning of hepatitis C virus genome from a single Japanese  
 RT carrier: sequence variation within the same individual and among  
 RT infected individuals.\*;  
 RL Virus Res. 23:39-53(1992).  
 CC -1- FUNCTION: THE SMALL PROTEINS NS2A, NS2B, NS4A AND NS4B ARE  
 CC HYDROPHOBIC, SUGGESTING A POSSIBLE MEMBRANE-RELATED FUNCTION.  
 CC NS3 AND NS5 MAY PLAY A ROLE IN THE VIRAL RNA REPLICATION.  
 CC -1- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral  
 CC precursor polypeptide, commonly with Asp or Glu in the P6  
 CC position, Cys or Thr in P1 and Ser or Ala in P1.  
 CC -1- CATALYTIC ACTIVITY: N nucleoside triphosphate -> N diphosphate +  
 CC (RNA)(N).  
 CC -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A  
 CC LIPID-PROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:  
 CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF  
 CC PROTEIN C AND MRNA.  
 CC -1- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY S29.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: D11168; BAA01943.1; -  
 DR PIR: A45573; A45573.  
 DR HSP: P26663; IJXP.  
 DR MEROPS: S29.001; -  
 DR InterPro: IPR001410; DEAD.  
 DR InterPro: IPR002531; HCV\_NS1.  
 DR InterPro: IPR002518; HCV\_NS2.  
 DR InterPro: IPR004109; HCV\_NS3.  
 DR InterPro: IPR000745; HCV\_NS4A.  
 DR InterPro: IPR001490; HCV\_NS4B.  
 DR InterPro: IPR002868; HCV\_NS5A.  
 DR InterPro: IPR002166; HCV\_RdRp.  
 DR InterPro: IPR002522; HCV\_capsid.  
 DR InterPro: IPR002521; HCV\_core.  
 DR InterPro: IPR002519; HCV\_env.  
 DR Pfam: PF00938; HCV\_RdRp.1.  
 DR Pfam: PF01001; HCV\_NS4b.1.  
 DR Pfam: PF01006; HCV\_NS4a.1.  
 DR Pfam: PF01506; HCV\_NS5a.1.  
 DR Pfam: PF01538; HCV\_NS2.1.  
 DR Pfam: PF01539; HCV\_NS2.1.  
 DR Pfam: PF01542; HCV\_core.1.  
 DR Pfam: PF01543; HCV\_capsid.1.  
 DR Pfam: PF01560; HCV\_NS1.1.  
 DR Pfam: PF02907; HCV\_NS3.1.  
 DR ProDom: PD186062; HCV\_NS1.1.  
 KW Polyprotein; Glycoprotein; Transferase; RNA-directed RNA polymerase;  
 KW Core protein; Coat protein; Envelope protein; Helicase; ATP-binding;  
 KW Transmembrane; Nonstructural protein; Hydrolase; Serine protease.

FT INT\_MET 1 1  
 FT CHAIN 1 115  
 FT CHAIN 116 191  
 FT CHAIN 192 383  
 FT CHAIN 384 729  
 FT CHAIN 730 1006  
 FT CHAIN 1007 1615  
 FT CHAIN 1616 1862  
 FT CHAIN 1863 2014  
 FT CHAIN 2014 3010  
 FT TRANSMEM 347 369  
 FT ACT\_SITE 1083 1083  
 FT ACT\_SITE 1107 1107  
 FT ACT\_SITE 1165 1165  
 FT NP\_BIND 1230 1317  
 FT SITE 1313 1319  
 FT CARBOHYD 196 196  
 FT CARBOHYD 209 209  
 FT CARBOHYD 234 234  
 FT CARBOHYD 250 250  
 FT CARBOHYD 305 305  
 FT CARBOHYD 417 417  
 FT CARBOHYD 423 423  
 FT CARBOHYD 430 430  
 FT CARBOHYD 448 448  
 FT CARBOHYD 532 532  
 FT CARBOHYD 540 540  
 FT CARBOHYD 556 556  
 FT CARBOHYD 576 576  
 FT CARBOHYD 623 623  
 FT CARBOHYD 645 645  
 FT CARBOHYD 2041 2041  
 FT CARBOHYD 2077 2077  
 FT CARBOHYD 2240 2240  
 FT CARBOHYD 2529 2529  
 FT CARBOHYD 2788 2788  
 SQ SEQUENCE 3010 AA; 326573 MW; 94A1C77435D642BB CRC64;  
  
 Alignment Scores:  
 Pred. No.: 5,76e-89  
 Score: 1073.00  
 Percent Similarity: 98.08%  
 Best Local Similarity: 94.71%  
 Query Match: 92.90%  
 DB: 1  
 Gaps: 0  
  
 US-09-899-303a-7 (1-633) x POLG\_HCVUT (1-3010)  
 QY 4 TTGGGTAAAGTCATGATACCTTACGTGGCGGCGGACCTTCATGGGCTACATTCG 63  
 DB 119 LeuGlyValIleAspThrLeuThrcysGlyPheAlaAspLeuMetGlyTyrIlePro 138  
 QY 64 CTGTCGGGGCGCCCGCTAGGGGGGTCGCCAGACCCCTGGCGGCGATGGCGTCCGGGTCTG 123  
 DB 139 LeuValGlyAlaProLeuGlyGlyAlaAlaValAlaLeuAlaHisGlyValAlaValIleu 158  
 QY 124 GAAGACGGCGGTGAATGCAACAGAGGGAATTTGCTGCTCTTCTATCTTCCTC 183  
 DB 159 GluAspGlyValAsnTyrAlaThrGlyAsnLeuProGlyCysSerPheSerIlePheLeu 178  
 QY 184 TTGGCTTACGTCTGCTGCTGACCATTCACCTTCGCTTATGAGGGTCCGACGTTCC 243  
 DB 179 LeuAlaLeuLeuSerCysLeuThrIleProAlaSerAlaTyrGlnValAlaGlnAlaSer 198  
 QY 244 GGGATGATCATGCTGCACGACGACTGCTCCCAACCAAGCATTTGTTAGAGCGCGGAC 303  
 DB 199 GlyLeuTyrHisValIleThrAsnAspCysSerAsnSerIleValIleTyrGlnAlaAlaGly 218  
 QY 304 ATGATCATGACACACCCCGGGTGGTCCCTGCTGGGAGGAACAACCTTCCCGCTGC 363  
 DB 219 MetIleMetHisThrProGlyCysValProCysValAlaArgGluAsnAsnAlaSerArgCys 238



```
FT CARBOHYD 417 417 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 423 423 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 430 430 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 448 448 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 532 532 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 540 540 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 556 556 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 576 576 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 623 623 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 645 645 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 2041 2041 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 2077 2077 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 2240 2240 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 2529 2529 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 2788 2788 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 3010 AA; 327189 MM; F842D5BECFDPD9C CRC64;

Alignment Scores:
Pred. No.: 8.59e-88 Length: 3010
Score: 1060.00 Matches: 194
Percent Similarity: 97.12% Conservative: 8
Best Local Similarity: 93.27% Mismatches: 6
Query Match: 91.77% Indels: 0
DB: 1 Gaps: 0

US-09-899-303a-7 (1-633) x POLG_HCVBK (1-3010)
OY 4 TTGGGTAAAGTCATGATACCCCTTACGCGCGCTTGGCGGACGTCATGGGGTACATTCG 63
Db 119 Lcuglylysvallieaprhleuthrcyselyphealaspblemeglytyrillepro 138
OY 64 CTCGTGCGCGCCCGCCCTAGGGGCTGTCGCGACCCCTGGCGCATGGCGGTCTGTG 123
Db 139 Lcualvalglialaproleuglyglialalalargalaleualahlsiglyvalargvalleu 158
OY 124 GAAGAGCGCGTGAACCTATGCAACAGGGAATTTGCTTCTTCTATCTCTCTC 183
Db 159 Gluaspelalyvalasnyralathrglyasneuproglcysserpheserilspheleu 178
OY 184 TTGGCTTACTGTCCTCTGTCGACCATTCACCTCCGCTTATGAGGAGCGGACGTGCC 243
Db 179 Lcualaleueneusercysleuthrthrpolaaserlatyrgluvalhlsasnyalser 198
OY 244 GGGATGACCATGTACAGAACGACTGCTCCAACTCAAGCATTTGTATGAGGACGCGAC 303
Db 199 Glyllyetrhlsvalthrasnaspysersasmlaserllevaltyrgluahlaalasp 218
OY 304 ATGATCATGCACACCCCGGGTGGCTGCTGCTGGGGAACAACCTCTCCGCTGC 363
Db 219 LcualleuthsthrproglcysvalpocysvalargglugljasnserSerArgcys 238
OY 364 TTGGGTAGCGCTGACCCCGCCAGCTGAGTGAAGCGACGCGCCGACATGACGACATA 423
Db 239 Trrvalalaleuthrprothrlleualalalargsnvalthrlleprothlthille 258
OY 424 CGAGCGCAGCTCGATTGCTGCTGGCGGCTCTTCTGCTTCGCTATGACTGCGGG 483
Db 259 Argtrghlsvalaspheuleuvalglialalalaphcysserlatametyvalgly 278
OY 484 GATCTCTGGGATCTGCTCTCTGCTGCCAGCTGTTACACATCTGCTCGCGCGCAT 543
Db 279 AspencysgliserValpheleuvalserclnleuphetrPheserProatrgarhls 298
OY 544 GAGAGCGTGAAGCATTCATTCATTCATTCATTCATTCATTCATTCATTCATTCAT 603
Db 299 valthrlleuclnaspccysasnccysserilleytrproolghlsvalserglyhlsargmet 318
OY 604 GCTTGGATATGATGATGAACCTGG 627
Db 319 AlatrpspmetmetmetasntTP 326

RESULT 6
POLG_HCVBK
```

```
ID POLG_HCVBK STANDARD; PRT; 520 AA.
AC 001403;
DT 01-JUL-1993 (Rel. 26, Created)
DT 01-JUL-1993 (Rel. 26, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Genome polyprotein [Contains: Capsid protein C (core protein) (P22);
DE Envelope glycoprotein E1 (GP32) (GP35); Envelope glycoprotein E2
DE (GP68) (GP70) (NS1)] (Fragment).
OS Hepatitis C virus (isolate HCV-KP) (HCV).
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_Taxid=31644;
RN [1]
RP MEDLINE=93019030; PubMed=1383400;
RA Abe K., Inchauspe G., Fujisawa K.;
RT "Genomic characterization and mutation rate of hepatitis C virus
RT isolated from a patient who contracted hepatitis during an epidemic
RT of non-A, non-B hepatitis in Japan."
RL J. Gen. Virol. 73:2725-2729(1992).
CC -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
CC PROTEIN C AND RNA.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: D10687; BAA01529.1; -.
DR PIR: J01925; J01925.
DR InterPro: IPR002531; HCV_NSI.
DR InterPro: IPR002522; HCV_capsid.
DR InterPro: IPR002521; HCV_core.
DR InterPro: IPR002519; HCV_env.
DR Pfam: PF01539; HCV_env; 1.
DR Pfam: PF01542; HCV_core; 1.
DR Pfam: PF01543; HCV_capsid; 1.
DR Pfam: PF01560; HCV_NSI; 1.
DR ProDom: PD186062; HCV_NSI; 1.
DR Transmembrane: Glycoprotein: Coat protein: Envelope protein;
KW Polypeptide; Glycoprotein; Nonstructural protein.
FT INIT_MET 1 1
FT CHAIN 1 115
FT CHAIN 116 191
FT CHAIN 192 383
FT CHAIN 384 >520
FT TRANSMEM 347 369
FT CARBOHYD 196 196
FT CARBOHYD 209 209
FT CARBOHYD 233 233
FT CARBOHYD 234 234
FT CARBOHYD 305 305
FT CARBOHYD 418 418
FT CARBOHYD 424 424
FT CARBOHYD 431 431
FT CARBOHYD 449 449
FT NON_TER 520 520
SQ SEQUENCE 520 AA; 56476 MM; 1D2BD0A6F27349B CRC64;

Alignment Scores:
Pred. No.: 1.24e-87 Length: 520
Score: 1057.00 Matches: 195
Percent Similarity: 96.63% Conservative: 6
Best Local Similarity: 93.75% Mismatches: 7
Query Match: 91.52% Indels: 0
DB: 1 Gaps: 0
```

US-09-899-303a-7 (1-633) x POLG\_HCVH4 (1-520)

QY 4 TTGGGTAAGTCATGATACCTTACGTCGGGCTTCCGACCTCATGGGGTACATTCCG 63

DB 119 LeuGlyLysValIleAspPhrIleuthrCysGlyPheAlaAspIleuMetGlyTyrIlePro 138

QY 64 CTCGTCGGGGCCCCCTAGGGGCTGCTGCCAAGAGCCCTGGCGATGGCGTCCGGTTCG 123

DB 139 LeuValGlyAlaProIleuLysGlyAlaSerArgAlaLeuAlaHisGlyValArgValLeu 158

QY 124 GAAGAGCGGCTGAATCATGCAACAGGAAATTGCTGTTGCTTCTTCTATCTCTC 183

DB 159 GluAspGlyValAsnTyrAlaThrGlyAsnLeuProGlyCysSerPheSerIlePheLeu 178

QY 184 TTGGCTTTACTGTCGTGTGACCAATTCAGCTCCGCTTATGAGGTGGCAAGCTGTCC 243

DB 179 SerAlaLeuMetSerCysLeuThrThrProAlaSerAlaTyrGlyValArgAsnValSer 198

QY 244 GGGATGACCATGTCTCGAAGACGCTGCAACTCAAGATGTTGTATGAGGACGGGAC 303

DB 199 GlyIleTyrHisValThrAsnAspCysSerAsnSerIleAlaTyrGlyAlaAlaGly 218

QY 304 ATGATCATGCACACCCCGGCTGCGCTGCGCTGCGGAGAACACTTTCGCGCTGC 363

DB 219 MetIleMetHisThrProGlyCysValProCysValArgGlyAsnAsnSerAlaGlyCys 238

QY 364 TTGGTAAGCCTCACCCCGGCTGCGCTGCGCTGCGGAGAACACTTTCGCGCTGC 423

DB 239 TrpValAlaLeuThrProThrIleuAlaAlaArgAsnAlaSerValProThrThrIle 258

QY 424 GCAGCGGACGTCGATGTCGTCGTCGGGGGCTGCTTCTGTTCCCTATGTCAGTGGG 483

DB 259 ArgAlaGHisValAspLeuValGlyAlaAlaThrLeuCysSerAlaMetTyrValGly 278

QY 484 GATCTCTCGGAGATCTGCTTCTCTCTGCTCCAGCTGTCACCATCTGCGCTCGCGGAC 543

DB 279 AspLeuCysGlySerValPheLeuValSerGlnLeuPheThrPheSerProAlaGlyTyr 298

QY 544 GAGACGCTGACGACTGCAATTGCTCAATCTATCCGGCCACATAAGGTCACCGTATG 603

DB 299 GluThrValGlnAspCysAsnSerIleTyrProGlyHisValSerGlyHisArgMet 318

QY 604 GCTTGGGATATGATGATGAACGTG 627

DB 319 AlATTPASpMetMetMetAsnTP 326

RESULT 7

POLG\_HCVH4 ID STANDARD: PRT: 520 AA.

AC 001404:

DB 01-JUL-1993 (Rel. 26, Created)

DB 01-JUL-1993 (Rel. 26, Last sequence update)

DB 16-OCT-2001 (Rel. 40, Last annotation update)

DB Genome polyprotein (contains: Capsid protein C (Core protein) (P22); Envelope glycoprotein E1 (GP32) (GP35); Envelope glycoprotein E2 (GP68) (GP70) (NS1) (Fragment)).

DB Hepatitis C virus (isolate HCV-4/6) (HCV).

DB Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae; Hepacivirus.

DB NCBI\_TaxID-31643:

DB [1]

DB SEQUENCE FROM N.A.

DB MEDLINE-93019030; PubMed-1383400;

DB Abe K., Inchausti G., Fujisawa K.;

DB "Genomic characterization and mutation rate of hepatitis C virus isolated from a patient who contracted hepatitis during an epidemic of non-A, non-B hepatitis in Japan."

DB J. Gen. Virol. 73:2725-2729(1992).

DB -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS: PROTEIN C AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF PROTEIN C AND RNA.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - CC the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way CC modified and this statement is not removed. Usage by and for commercial CC entities requires a license agreement (See <http://www.isb-sdb.ch/announce/> or send an email to [license@isb-sdb.ch](mailto:license@isb-sdb.ch)).

CC -----

DB EMBL: D10688; BAA01530.1; -

DB InterPro: IPR002531; HCV\_NS1.

DB InterPro: IPR002522; HCV\_capsid.

DB InterPro: IPR002521; HCV\_core.

DB InterPro: IPR002519; HCV\_env.

DB Pfam: PF01539; HCV\_env; 1.

DB Pfam: PF01542; HCV\_core; 1.

DB Pfam: PF01543; HCV\_capsid; 1.

DB Pfam: PF01560; HCV\_NS1; 1.

DB ProDom: PD186062; HCV\_NS1; 1.

DB Polyprotein; glycoprotein; Coat protein; Envelope protein; Transmembrane; Nonstructural protein.

DB INIT\_MET 1

DB CHAIN 1

DB CHAIN 115

DB CHAIN 191

DB CHAIN 192

DB CHAIN 383

DB CHAIN 384

DB CHAIN >520

DB TRANSMEM 369

DB CARBOHYD 196

DB CARBOHYD 196

DB CARBOHYD 209

DB CARBOHYD 209

DB CARBOHYD 234

DB CARBOHYD 305

DB CARBOHYD 305

DB CARBOHYD 418

DB CARBOHYD 424

DB CARBOHYD 431

DB CARBOHYD 449

DB CARBOHYD 520

DB NON\_TER 520

DB SEQUENCE 520 AA; 56499 MW; AA135246CF20D525 CRC64;

Alignment Scores:

Pred. No.: 1.85e-86

Score: 1044.00

Percent Similarity: 96.15%

Best Local Similarity: 92.31%

Query Match: 90.39%

DB: 1

Gaps: 0

US-09-899-303a-7 (1-633) x POLG\_HCVH4 (1-520)

QY 4 TTGGGTAAGTCATGATACCTTACGTCGGGCTTCCGACCTCATGGGGTACATTCCG 63

DB 119 LeuGlyLysValIleAspPhrIleuthrCysGlyPheAlaAspIleuMetGlyTyrIlePro 138

QY 64 CTCGTCGGGGCCCCCTAGGGGCTGCTGCCAAGAGCCCTGGCGATGGCGTCCGGTTCG 123

DB 139 LeuValGlyAlaProIleuLysGlyAlaSerArgAlaLeuAlaHisGlyValArgValLeu 158

QY 124 GAAGAGCGGCTGAATCATGCAACAGGAAATTGCTGTTGCTTCTTCTATCTCTC 183

DB 159 GluAspGlyValAsnTyrAlaThrGlyAsnLeuProGlyCysSerPheSerIlePheLeu 178

QY 184 TTGGCTTTACTGTCGTGTGACCAATTCAGCTCCGCTTATGAGGTGGCAAGCTGTCC 243

DB 179 SerAlaLeuMetSerCysLeuThrThrProAlaSerAlaTyrGlyValArgAsnValSer 198

QY 244 GGGATGACCATGTCTCGAAGACGCTGCAACTCAAGATGTTGTATGAGGACGGGAC 303

DB 199 GlyIleTyrHisValThrAsnAspCysSerAsnSerIleAlaTyrGlyAlaAlaGly 218

QY 304 ATGATCATGCACACCCCGGCTGCGCTGCGCTGCGGAGAACACTTTCGCGCTGC 363

DB 219 MetIleMetHisThrProGlyCysValProCysValArgGlyAsnAsnSerAlaGlyCys 238

QY 364 TTGGTAAGCCTCACCCCGGCTGCGCTGCGCTGCGGAGAACACTTTCGCGCTGC 423

```

Db      239  TrrpAlaIaLeuThrProThrLeuAlaIaArgAsnAlaSerValProThrThrThrIle 258
Qy      424  CGAGCCACGACGATTTGCTGTTGGGGCGGCTTTCTGTCGGATGATGAGG 483
Db      259  ArgArgHisValAspLeuLeuValGlyAlaIaIaThrLeuGlySerAlaMetGlyValGly 278
Qy      484  GATCTGTCGGGATGTCCTTCCTGCTCCAGCTGTTACCATCTGCGCTCGCGGACAT 543
Db      279  AspLeuGlySerValPheLeuValSerGlnLeuIleThrPheSerProArgArgTyr 298
Qy      544  GAGACGGTGCAGACTGCAATTCATATCCCGGCGCAGAACAGGTCACCGTATG 603
Db      299  GluThrValGlnAspCysAsnCysSerLeuTyrProGlyHisValSerGlyHisArgMet 318
Qy      604  GCTTGGATATGATGATCACTGG 627
Db      319  AlaTrpAspMetMetMetAsnTrp 326

RESULT 8
ID POLG_HCVH STANDARD; PRT; 3011 AA.
AC P27958;
DT 01-AUG-1992 (Rel. 23, Created)
DT 01-AUG-1992 (Rel. 23, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Genome polypeptide [Contains: Capsid protein C (core protein) (P22);
DE Envelope glycoprotein E1 (GP32) (GP35): Envelope glycoprotein E2
DE (GP68) (GP70) (NS1): Protein p7; Nonstructural protein NS2 (P21)
DE (EC 3.4.22.-); Protease/helicase NS3 (P70) (Hepatitisin)
DE (EC 3.4.21.98): Nonstructural protein NS4A (P4): Nonstructural protein
DE NS4B (P66) (P70) (RNA-directed RNA polymerase) (EC 2.7.7.48)].
OS Hepatitis C virus (isolate H) (HCV).
OC Viruses: ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11108;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92052256; PubMed=1658800;
RA Inchauspe G., Zebadee S., Lee D.H.H., Sugitani M., Nasoff M.,
RA Prince A.M.:
RT "Genomic structure of the human prototype strain H of hepatitis C
RT virus: comparison with American and Japanese isolates."
RL Proc. Natl. Acad. Sci. U.S.A. 88:10292-10296(1991).
RN [2]
RP X-RAY CRYSTALLOGRAPHY (2.1 ANGSTROMS) OF 1207-1657.
RX MEDLINE=97331322; PubMed=9187654.
RA Yeo N., Hesson T., Cable M., Hong Z., Kwong A.D., Le H.V., Weber P.C.:
RT "Structure of the hepatitis C virus RNA helicase domain."
RL Nat. Struct. Biol. 4:463-467(1997).
RN [3]
RP X-RAY CRYSTALLOGRAPHY (2.2 ANGSTROMS) OF 1192-1657.
RX MEDLINE=98154321; PubMed=9493270.
RA Kim J.L., Morgenstern K.A., Grifflith J.P., Dwyer M.D., Thomson J.A.,
RA Murcho M.A., Lin C., Caron P.R.:
RT "Hepatitis C virus NS3 RNA helicase domain with a bound
RT oligonucleotide: the crystal structure provides insights into the mode
RT of unwinding."
RL Structure 6:89-100(1998).
CC -1- FUNCTION: PROTEASE NS2 IS RESPONSIBLE FOR THE CLEAVAGE OF NS2-NS3.
CC -1- FUNCTION: PROTEASE NS3 IS RESPONSIBLE FOR THE CLEAVAGE OF
CC NS3-NS4A, NS4A-NS4B, NS4B-NS5A AND NS5A-NS5B.
CC -1- FUNCTION: NS4A FORMS A COMPLEX WITH NS3 AND IS ESSENTIAL FOR THE
CC ACTIVATION OF NS3.
CC -1- FUNCTION: NS5A SEEMS TO HAVE A TRANSCRIPTIONAL ACTIVATORY ROLE.
CC -1- FUNCTION: NS5B IS A RNA-DEPENDENT RNA POLYMERASE THAT PLAYS AN
CC ESSENTIAL ROLE IN THE VIRUS REPLICATION.
CC -1- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral
CC precursor polyprotein, commonly with Asp or Glu in the P6
CC position, Cys or Thr in P1 and Ser or Ala in P1'.
CC -1- CATALYTIC ACTIVITY: N nucleoside triphosphate -> N diphosphate +
CC [RNA](N).
```

```

CC -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS: E1
CC AND E2. THE NUCLEOCAPSID IS A COMPLEX OF PROTEIN C AND RNA.
CC -1- PTM: THE STRUCTURAL PROTEINS C, E1 AND E2 ARE PRODUCED BY
CC PROTEOLYTIC PROCESSING BY THE HOST SIGNAL PEPTIDASES.
CC -1- SIMILARITY: THE NS2 PROTEASE BELONGS TO PEPTIDASE FAMILY C18.
CC -1- SIMILARITY: THE NS3 PROTEASE BELONGS TO PEPTIDASE FAMILY S29.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: M67463; AAA45534.1; -.
DR PIR: A36814; GNMVCH.
DR PDB: 1HEI; 25-NOV-98.
DR PDB: 1AIV; 16-FEB-99.
DR MEROPS: S29.001; -.
DR MEROPS: U39.001; -.
DR TRANSFAC: T04155; -.
DR InterPro: IPR001410; DEAD.
DR InterPro: IPR002531; HCV_NS1.
DR InterPro: IPR002518; HCV_NS2.
DR InterPro: IPR004109; HCV_NS3.
DR InterPro: IPR000745; HCV_NS4A.
DR InterPro: IPR001490; HCV_NS4B.
DR InterPro: IPR002868; HCV_NS5A.
DR InterPro: IPR002166; HCV_RdRP.
DR InterPro: IPR002522; HCV_capsid.
DR InterPro: IPR002521; HCV_core.
DR InterPro: IPR002519; HCV_env.
DR InterPro: IPR001650; Helicase_C.
DR Pfam: PF00271; helicase_C; 1.
DR Pfam: PF00998; HCV_RdRP; 1.
DR Pfam: PF01001; HCV_NS4B; 1.
DR Pfam: PF01006; HCV_NS4A; 1.
DR Pfam: PF01506; HCV_NS5A; 1.
DR Pfam: PF01538; HCV_NS2; 1.
DR Pfam: PF01539; HCV_env; 1.
DR Pfam: PF01542; HCV_core; 1.
DR Pfam: PF01543; HCV_capsid; 1.
DR Pfam: PF01560; HCV_NS1; 1.
DR Pfam: PF02907; HCV_NS3; 1.
DR ProDom: PD186062; HCV_NS1; 1.
DR Core protein; Glycoprotein; Transferase: RNA-directed RNA polymerase;
DR Core protein; Coat protein; Envelope protein; Helicase; ATP-binding;
DR Transmembrane; Nonstructural protein; Hydrolase; Serine protease;
DR 3D-structure.
DR INT_MET 1 1
FT CHAIN 1 191
FT CHAIN 192 383
FT CHAIN 384 746
FT CHAIN 747 809
FT CHAIN 810 1026
FT CHAIN 1027 1657
FT CHAIN 1658 1711
FT CHAIN 1712 1972
FT CHAIN 1973 2420
FT CHAIN 2421 3011
FT CHAIN 3011 369
FT TRANSMEM 347 369
FT ACT_SITE 1083 1083
FT ACT_SITE 1107 1107
FT ACT_SITE 1165 1165
FT NP_BIND 1230 1237
FT SITE 1316 1319
FT CARBOHYD 136 196
FT CARBOHYD 209 209
FT CARBOHYD 234 234
FT CARBOHYD 305 305

REMOVED FROM CAPSID PROTEIN C BY THE
CELLULAR AMINOPEPTIDASE.
CAPSID PROTEIN C.
ENVELOPE GLYCOPROTEIN E1.
ENVELOPE GLYCOPROTEIN E2.
PROTEIN P7.
NONSTRUCTURAL PROTEIN NS2.
PROTEASE/HELICASE NS3.
NONSTRUCTURAL PROTEIN NS4A.
NONSTRUCTURAL PROTEIN NS4B.
NONSTRUCTURAL PROTEIN NS5A.
NONSTRUCTURAL PROTEIN NS5B.
POTENTIAL.
CHARGE RELAY SYSTEM (BY SIMILARITY).
CHARGE RELAY SYSTEM (BY SIMILARITY).
ATP (POTENTIAL).
DECH BOX.
N-LINKED (GLCNAC. . .) (POTENTIAL).
N-LINKED (GLCNAC. . .) (POTENTIAL).
N-LINKED (GLCNAC. . .) (POTENTIAL).
N-LINKED (GLCNAC. . .) (POTENTIAL).
N-LINKED (GLCNAC. . .) (POTENTIAL).
```

SO	SEQUENCE	3011 AA:	527142 MM:	772CB8239C0B94753 CRO64:
FT	CARBOHYD	417	417	(POTENTIAL)
FT	CARBOHYD	423	423	N-LINKED (GLCNAC. . .) (POTENTIAL)
FT	CARBOHYD	430	430	N-LINKED (GLCNAC. . .) (POTENTIAL)
FT	CARBOHYD	448	448	N-LINKED (GLCNAC. . .) (POTENTIAL)
FT	CARBOHYD	476	476	N-LINKED (GLCNAC. . .) (POTENTIAL)
FT	CARBOHYD	532	532	N-LINKED (GLCNAC. . .) (POTENTIAL)
FT	CARBOHYD	540	540	N-LINKED (GLCNAC. . .) (POTENTIAL)
FT	CARBOHYD	556	556	N-LINKED (GLCNAC. . .) (POTENTIAL)
FT	CARBOHYD	576	576	N-LINKED (GLCNAC. . .) (POTENTIAL)
FT	CARBOHYD	623	623	N-LINKED (GLCNAC. . .) (POTENTIAL)
FT	CARBOHYD	645	645	N-LINKED (GLCNAC. . .) (POTENTIAL)

**Alignment Scores:**

Pred. No.:	6.3e-62	Length:	301
Score:	995.00	Matches:	178
Percent Similarity:	92.798	Conservative:	15
Best Local Similarity:	85.584	Mismatches:	15
Query Match:	86.154	Indels:	0
DB:	1	Gaps:	0

US-09-899-303A-7 (1-633) x POLG\_HCVH (1-3011)

[illegible]

## RESULT 9

ID	POLG_HCV1	STANDARD;	PRT;	3011 AA.
AC	P26664;			
DT	01-AUG-1992	(Rel. 23, Created)		
DT	01-AUG-1992	(Rel. 23, Last sequence update)		

DT 15-JUN-2002 (Rel. 41, Last annotation update)  
DE Envelope glycoprotein [contains: Capsid protein C (Core protein) (P22);  
DE Envelope glycoprotein E1 (GP33); Envelope glycoprotein E2  
DE (GP68) (GP70) (NS1); Protein P7; Nonstructural protein NS2 (P21)  
DE (EC 3.4.22.-); Protease/helicase NS3 (P70) (Hepacivirin)  
DE (EC 3.4.21.99); Nonstructural protein NS4A (P4); Nonstructural protein  
DE NS4B (P27); Nonstructural protein NS5A (P56); Nonstructural protein  
DE NS5B (P66) (P70) (RNA-directed RNA polymerase) (EC 2.7.7.48)].  
OS Hepatitis C virus (isolate 1) (HCV)  
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;  
OC Hepacivirus.  
OC NCBI\_TaxId=11104;  
OX [1]  
RN  
RP SEQUENCE FROM N.A.  
RX MEDLINE=91172826; PubMed=1848704;  
RA Choo O.-L., Richman K.H., Han J.H., Berger K., Lee C., Dong C.,  
RA Gallegos C., Colt D., Medina-Selby A., Barr P.J., Weiner A.J.,  
RA Bradley D.W., Kuo G., Houghton M.;  
RT "Genetic organisation and diversity of the hepatitis C virus.";  
RL Proc. Natl. Acad. Sci. U.S.A. 88:2451-2455(1991).  
CC -1- FUNCTION: THE SMALL PROTEINS NS2A, NS2B, NS4A AND NS4B ARE  
CC NS3 AND NS5 MAY PLAY A ROLE IN THE VIRAL RNA REPLICATION.  
CC HYDROPHOBIC, SUGGESTING A POSSIBLE MEMBRANE-RELATED FUNCTION.  
CC -1- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral  
CC precursor polypeptide, commonly with Asp or Glu in the P6  
CC position, Cys or Thr in P1 and Ser or Ala in P1'.  
CC -1- CATALYTIC ACTIVITY: N nucleoside triphosphate = N diphosphate +  
CC (RNA)(N).  
CC -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A  
CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF  
CC PROTEIN C AND MNA.  
CC -1- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY S29.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; M62321; AAA5676.1; -.  
DR pir; A39166; GNMVCA.  
DR HSSP; P27958; 1A1V.  
DR MEROPS; S29.001; -.  
DR MEROPS; U39.001; -.  
DR InterPro: IPR001410; DEAD.  
DR InterPro: IPR002531; HCV\_NS1.  
DR InterPro: IPR002518; HCV\_NS2.  
DR InterPro: IPR004109; HCV\_NS3.  
DR InterPro: IPR000745; HCV\_NS4A.  
DR InterPro: IPR001490; HCV\_NS4B.  
DR InterPro: IPR002868; HCV\_NS5A.  
DR InterPro: IPR002166; HCV\_NS5B.  
DR InterPro: IPR002522; HCV\_RdRp.  
DR InterPro: IPR002521; HCV\_core.  
DR InterPro: IPR002519; HCV\_env.  
DR InterPro: IPR001650; Helicase\_C.  
DR Pfam; PF00271; Helicase\_C; 1.  
DR Pfam; PF00998; HCV\_RdRp; 1.  
DR Pfam; PF01001; HCV\_NS4B; 1.  
DR Pfam; PF01006; HCV\_NS4A; 1.  
DR Pfam; PF01506; HCV\_NS5A; 1.  
DR Pfam; PF01538; HCV\_NS2; 1.  
DR Pfam; PF01539; HCV\_env; 1.  
DR Pfam; PF01543; HCV\_core; 1.  
DR Pfam; PF01543; HCV\_capsid; 1.  
DR Pfam; PF01560; HCV\_NS1; 1.  
DR Pfam; PF02907; HCV\_NS3; 1.  
DR ProDom; PD186062; HCV\_NS1; 1.  
DR PolyProtein; Glycoprotein; Transferase; RNA-directed RNA polymerase;  
KW Core protein; Coat protein; Envelope protein; Helicase; ATP-binding;

KM Transmembrane; Nonstructural protein; Hydrolase; Serine protease.  
 FT INT\_MET 1 1  
 FT CHAIN 1 115  
 FT CHAIN 116 191  
 FT CHAIN 192 383  
 FT CHAIN 384 729  
 FT CHAIN 730 1006  
 FT CHAIN 1007 1615  
 FT CHAIN 1616 1862  
 FT CHAIN 1863 2013  
 FT CHAIN 2014 3011  
 FT TRANSMEM 347 369  
 FT ACT\_SITE 1083 1083  
 FT ACT\_SITE 1107 1107  
 FT ACT\_SITE 1165 1165  
 FT NE\_BIND 1230 1237  
 FT SITE 1316 1319  
 FT CARBOHYD 196 196  
 FT CARBOHYD 209 209  
 FT CARBOHYD 234 234  
 FT CARBOHYD 305 305  
 FT CARBOHYD 417 417  
 FT CARBOHYD 423 423  
 FT CARBOHYD 430 430  
 FT CARBOHYD 448 448  
 FT CARBOHYD 476 476  
 FT CARBOHYD 532 532  
 FT CARBOHYD 540 540  
 FT CARBOHYD 556 556  
 FT CARBOHYD 576 576  
 FT CARBOHYD 623 623  
 FT CARBOHYD 645 645  
 FT CARBOHYD 2041 2041  
 FT CARBOHYD 2077 2077  
 FT CARBOHYD 2240 2240  
 FT CARBOHYD 2364 2364  
 FT CARBOHYD 2789 2789  
 SQ SEQUENCE 3011 AA; 327197 MW; 65F8C9447FCE5AF9 CRC64;

Alignment Scores:  
 Pred. No.: 1,18e-81 Length: 3011  
 Score: 992.00 Matches: 178  
 Percent Similarity: 92.31% Conservative: 14  
 Best Local Similarity: 85.58% Mismatches: 16  
 Query Match: 85.89% Indels: 0  
 DB: 1 Gaps: 0

US-09-899-303a-7 (1-633) x POLG\_HCV1 (1-3011)

QY 4 TTGGGTAAAGCTCATGATACCTTACCTGCGGCTTCCGACACCTCATGAGGATACATTCG 63  
 DB 119 Lengllyslvalllleasphrlleuthrlycsglyphenalaspheumetglttyrlllepro 138  
 QY 64 CTGCTGGGCGCCCGCTAGGGGCTGCGACAGCCCTGGCGCATGGCGCTCGGGTCTG 123  
 DB 139 lenuvalgllyalproleugllylalaalarglaleuallahscglyvalargvalleu 158  
 QY 124 GAAGAGCGCGTGAACATGACAGGGAATTTGGCTTGCCTTCTCATCTCTCCTC 183  
 DB 159 gluspelvalasnylralnhrlyalshndeuprollycysserpsestlephelu 178  
 QY 184 TTGGCTTACTGCTCTCTGACCATTCAGCTTCCGCTTATGAGTGGCCCAAGTCTCC 243  
 DB 179 lenualeuenseuercysleuthrvalproalaserlalyrglnvalargasnerthr 198  
 QY 244 GGGATGACATGTCACGAAGACGACTGCTCAACTCAAGCATTTGATGAGGACGGAC 303  
 DB 199 glyleuylrhlshvaltrhranaspqysproasnserserlevallyrglnalalaasp 218  
 QY 304 ATATGATGACACACCCCGGGTGGCTGCTGCTGGGGAACACATCTGCTCCGGTGC 363  
 DB 219 Alalleuulshthrprogllycysvalprocysvalarglgluglnalaseratgcs 238

QY 364 TTGGTACGCTCACCCCGCTGCGACGTAGAGACCGACGTCCTCCACTACGACATA 423  
 DB 239 TrpAlalmetethrprothrvalalathrArgaspllylsleuProalathrGlnLeu 258  
 QY 424 CGAGCCGACGTGATTTGCTGCTGGGGCGGCTTTCGTCGGTATGACGTGGG 483  
 DB 259 ArgArghlshlleaspleuauValgllySerAlathrdeucysSerAlaleuTyValglly 278  
 QY 484 GATCTGGCGGAGTCTGCTCTGCTGCTCCAGCTGTCACCATCTGCGCTCGCGGCAT 543  
 DB 279 AspleuysgllyservalpheleuValgllyGlnleuThrphrPheSerProAlaGlnHis 298  
 QY 544 GAGACGGTGCAGACACTGCAATTGCTCATCTATCCGCGCACATACAGGTACACCTATG 603  
 DB 299 TrpThrthrGlnGlycysAsnGlySerIleTyThrProgllyshlslethrGlnHisArgMet 318  
 QY 604 GCTTGGATATGATGATGAACCTGG 627  
 DB 319 AlatrpsamethmetMetasnrtr 326

RESULT 10  
 POLG\_HCVTH STANDARD; PRT; 321 AA.  
 ID P27957;  
 AC 01-AUG-1992 (Rel. 23, Created)  
 DT 01-AUG-1992 (Rel. 23, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Genome polyprotein [contains: Matrix protein (Envelope protein M);  
 DE Major envelope protein E; Nonstructural protein NS1 (Fragment).  
 OS Hepatitis C virus (isolate TH) (HCV).  
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;  
 OC Hepacivirus.  
 NCBI\_TaxID=11117;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE-91112009; PubMed-1846505;  
 RA Weiner A.J., Brauer M.J., Rosenblatt J., Richman K.H., Tung J.,  
 RA Crawford K., Bonino F., Saracco G., Choo Q.-L., Houghton M.,  
 RA Han J.H.;  
 RT "Variable and hypervariable domains are found in the regions of HCV  
 RT corresponding to the flavivirus envelope and NS1 proteins and the  
 RT pestivirus envelope glycoproteins.";  
 RL Virology 180:842-848(1991).  
 CC -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A  
 CC LIPID-PROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:  
 CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF  
 CC PROTEIN C AND RNA.  
 CC -----  
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL: X51314; CAA37294.1; -  
 CC InterPro: IPR002521; HCV\_NS1.  
 CC InterPro: IPR002523; HCV\_core.  
 CC InterPro: IPR002519; HCV\_env.  
 CC Pfam: PF01539; HCV\_env; 1.  
 CC Pfam: PF01542; HCV\_core; 1.  
 CC Pfam: PF01560; HCV\_NS1; 1.  
 CC ProDom: PD186062; HCV\_NS1; 1.  
 KW Polyprotein; Glycoprotein; Coat protein; Envelope protein;  
 KW Transmembrane; Nonstructural protein.  
 FT NON\_TER 1 1  
 FT CHAIN <1 75  
 FT CHAIN 76 267  
 FT CHAIN 268 >321  
 FT CARBOHYD 80 80  
 FT CARBOHYD 93 93  
 N-TERMINAL  
 MATRIX PROTEIN (POTENTIAL).  
 MAJOR ENVELOPE PROTEIN E (POTENTIAL).  
 NONSTRUCTURAL PROTEIN NS1/E2 (POTENTIAL).  
 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 118 118 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 189 189 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 301 301 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 307 307 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 314 314 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT NON\_TER 321 321  
 SQ SEQUENCE 321 AA; 34074 MW; B2EB83F521C3B520 CRC64;

Alignment Scores:  
 Pred. No.: 1,966-81 Length: 321  
 Score: 988.00 Matches: 176  
 Percent Similarity: 92.31% Conservative: 16  
 Best Local Similarity: 84.62% Mismatches: 16  
 Query Match: 85.54% Indels: 0  
 Gaps: 0

US-09-899-303a-7 (1-633) x POLG\_HCVH8 (1-321)

QY 4 TTGGGTAAGGTATGATGATACCTTACGTGCGGCTTGGCCGACCTCATGGGGTACATTCCG 63  
 DB 3 LeuGlyLVysValIleAspThrLeuThrCysGlyPheAlaAspLeuMetGlyTyrIlePro 22  
 QY 64 CTCGTGGGCCCCCTAGGGGTGCTGCCAGAGCCCTGGCATGGCGTGGGTTCTG 123  
 DB 23 LeuValGlyAlaProLeuGlyGlyAlaAlaArgAlaLeuAlaHisGlyValaArgValLeu 42  
 QY 124 GAAGACGCGGTGAATATGCAACAGGGAATTGCTGTTCTTCTATCTTCCTC 183  
 DB 43 GLUAspGlyValaIleAsnTyrIleThrGlyAsnLeuProGlyCysSerPheSerLeuPheLeu 62  
 QY 184 TTGGCTTACGTGCTGCTGACCATTCGACGCTTCCGCTATGAGTGGCGCAACGTCTCC 243  
 DB 63 LeuAlaLeuLeuSerCysLeuThrValProAlaSerAlaTyrGlnValaArgAsnSerThr 82  
 QY 244 GGGATTCACCATGATGCAAGACGACTGCTCAACTCAAGCATTTGTATAGAGCAGCGAG 303  
 DB 83 GlyLeuTyrHisValIleThrAsnAspCysProAlaSerSerIleValTyrGlnAlaAlaAsp 102  
 QY 304 ATGATCATGCAACACCCCGGGTGCCTGCTGCGGAGAACACTCTTCCGCTGC 363  
 DB 103 AlaIleLeuHisAlaProGlyCysValProCysValaArgGlnGlyAsnAlaSerAlaCys 122  
 QY 364 TGGGTAGCGCTACCCCGGCGGCTAGGAGCCAGCGCTCCCGCACTAGCAATA 423  
 DB 123 TrrValAlaMetSerThrProThrValAlaThrArgAspGlyArgLeuThrThrGlnLeu 142  
 QY 424 CSAGCGSACGATGATGCTGCTGCTGCGGCGGCTGCTTGTCTCCGCTATGACGTGGG 483  
 DB 143 ArgArgHisIleLeuSerLeuValaGlySerAlaIleThrLeuCysSerAlaLeuTyrValaGly 162  
 QY 484 GATCTGCGGATCTGTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 543  
 DB 163 AspLeuCysGlySerIlePheLeuValaGlyGlnLeuPheThrPheSerProAlaArgHis 182  
 QY 544 GAGAGCGTCGACGACGAAATGCTCAATCTATCCGCGACATACAGCTACCCCTAG 603  
 DB 183 TrrThrThrGlnGlyCysAsnCySerIleTyrProGlyHisIleThrGlyHisArgMet 202  
 QY 604 GCTTGGGATATGATGATGAATGG 627  
 DB 203 AlaTrrAspMetMetMetAsnTrr 210

RESULT 11

POLG\_HCVH8 STANDARD: PRT; 321 AA.  
 AC P27956;  
 DT 01-AUG-1992 (rel. 23. Created)  
 DT 01-AUG-1992 (rel. 23. Last sequence update)  
 DT 16-OCT-2001 (rel. 40. Last annotation update)  
 DE Genome polyprotein [contains: Matrix protein (Envelope protein M);  
 DE Major envelope protein E; Nonstructural protein NS1] (Fragment).  
 OS Hepatitis C virus (isolate HCV18) (HCV).  
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;

OC Hepacivirus.  
 OX NCBI\_TaxId=11110;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=91112009; PubMed=1846505;  
 RA Weiner A.J., Brauer M.J., Rosenblatt J., Richman K.H., Tung J.,  
 RA Crawford K., Bonino F., Saracco G., Choo Q.-L., Houghton M.,  
 RA Han J.H.;  
 RA \*Variable and hypervariable domains are found in the regions of HCV  
 RT corresponding to the flavivirus envelope and NS1 proteins and the  
 RL pestivirus envelope glycoproteins.";  
 CC Virology 180:842-848(1991).  
 CC -I- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A  
 CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:  
 CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF  
 CC PROTEIN C AND MRNA.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----

DR EMBL: X53131; CAA37291.1; -  
 DR InterPro: IPR002531; HCV\_NSI.  
 DR InterPro: IPR002521; HCV\_core.  
 DR InterPro: IPR002519; HCV\_env.  
 DR Pfam: PF01539; HCV\_env; 1.  
 DR Pfam: PF01542; HCV\_core; 1.  
 DR Pfam: PF01560; HCV\_NSI; 1.  
 DR ProDom: PD186062; HCV\_NSI; 1.  
 KW Polyprotein; Glycoprotein; Coat protein; Envelope protein;  
 KW Transmembrane; Nonstructural protein.

FT CHAIN 1 76  
 FT CHAIN 267 76  
 FT CHAIN 268 >321  
 FT CARBOHYD 80 80 MAJOR ENVELOPE PROTEIN E (POTENTIAL).  
 FT CARBOHYD 93 93 NONSTRUCTURAL PROTEIN NS1/E2 (POTENTIAL).  
 FT CARBOHYD 118 118 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 189 189 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 301 301 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 307 307 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 314 314 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT NON\_TER 321 321  
 SQ SEQUENCE 321 AA; 34238 MW; 2F5DE79F7C7845C8 CRC64;

Alignment Scores:  
 Pred. No.: 5,536-81 Length: 321  
 Score: 983.00 Matches: 177  
 Percent Similarity: 91.83% Conservative: 14  
 Best Local Similarity: 85.10% Mismatches: 17  
 Query Match: 85.11% Indels: 0  
 Gaps: 0

US-09-899-303a-7 (1-633) x POLG\_HCVH8 (1-321)

QY 4 TTGGGTAAGGTATGATGATACCTTACGTGCGGCTTGGCCGACCTCATGGGGTACATTCCG 63  
 DB 3 LeuGlyLVysValIleAspThrLeuThrCysGlyPheAlaAspLeuMetGlyTyrIlePro 22  
 QY 64 CTCGTGGGCCCCCTAGGGGTGCTGCCAGAGCCCTGGCATGGCGTGGGTTCTG 123  
 DB 23 LeuValGlyAlaProLeuGlyGlyAlaAlaArgAlaLeuAlaHisGlyValaArgValLeu 42  
 QY 124 GAAGACGCGGTGAATATGCAACAGGGAATTGCTGTTCTTCTATCTTCCTC 183  
 DB 43 GLUAspGlyValaIleAsnTyrIleThrGlyAsnLeuProGlyCysSerPheSerLeuPheLeu 62  
 QY 184 TTGGCTTACGTGCTGCTGACCATTCGACGCTTCCGCTATGAGTGGCGCAACGTCTCC 243

Db 63 LeuAlaLeuSerCysLeuThrValProAlaSerAlaHisGlnValArgAsnSerThr 82  
 QY 244 GGGATGACCATGTCACGACGACGACTGCTCCAGCTCAAGCATTTGTATGAGCAGCGGAC 303  
 Db 83 GlyLeuThrHisValThrAspAlaAspCysProAsnSerSerIleValTyrGluAlaAlaAsp 102  
 QY 304 ATGATTCATGCACACCCCGGCGTGGCCCTGCGGAGAACACTCTCCCGCGTGC 363  
 Db 103 AlaIleLeuHisThrProGlyCysValProCysValHisGluGlyValAlaSerArgCys 122  
 QY 364 TGGGTACCGCTGACCCCGACGCTGACGAGTATCCGCGGACGACGACGACGACGACATA 423  
 Db 123 TrpValAlaValThrProThrValAlaIleThrArgAspGlyLysLeuProThrThrGlnLeu 142  
 QY 424 CGAGCCGACGTCGATTTGCTGCTGGGCGGCTGCTTGTCTGCTGCTGATGACGTCGGG 483  
 Db 143 ArgArgHisIleLeuAspLeuValGlySerAlaThrLeuCysSerAlaLeuTyrValGly 162  
 QY 484 GATCTCTGCGGATCTGCTCTCTGCTGCTCCAGCTGTTCACCACTGCGCTGCGCGCAT 543  
 Db 163 AspLeuCysGlySerValAlaPheLeuValGlyGlnLeuPheThrPheSerProArgArgHis 182  
 QY 544 GAGACGGTGGACGACGTCGATTTGCTCAATGCTATCCGCGGACGACGACGTCACGCTATG 603  
 Db 183 TrpThrThrGlnGlyCysAsnCysSerIleTyrProGlyHisIleThrGlyHisArgMet 202  
 QY 604 GCTTGGATATGATGATGATGATGATG 627  
 Db 203 AlaTrpAspMetMetMetLeuTyr 210  
 RESULT 12  
 POLG\_HCVH7  
 ID POLG\_HCVH7 STANDARD: PRT: 309 AA.  
 AC P27955;  
 DT 01-AUG-1992 (Rel. 23, Created)  
 DT 01-AUG-1992 (Rel. 23, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Genome polypeptide [Contains: Matrix protein (Envelope protein M);  
 DE Major envelope protein E; Nonstructural protein NS1] (Fragment).  
 OS Hepatitis C virus (isolate HCT27) (HCV).  
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;  
 OC Hepacivirus.  
 OC NCBI\_TaxID=11109;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=91112009; PubMed=1846505;  
 RA Weiner A.J., Brauer M.J., Rosenblatt J., Richman K.H., Tung J.,  
 RA Crawford K., Bonino F., Saracco G., Choo Q.-L., Houghton M.,  
 RA Han J.H.;  
 RT "Variable and hypervariable domains are found in the regions of HCV  
 RT corresponding to the flavivirus envelope and NS1 proteins and the  
 RT pestivirus envelope glycoproteins.";  
 RL Virology 180:842-848(1991).  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC EMBL: X53133; CAA37293.1; -  
 DR InterPro: IPR002531; HCV\_NSI.  
 DR InterPro: IPR002521; HCV\_core.  
 DR InterPro: IPR002519; HCV\_env.  
 DR Pfam: PF01539; HCV\_env; 1.  
 DR Pfam: PF01542; HCV\_core; 1.  
 DR Pfam: PF01560; HCV\_NSI; 1.  
 DR ProDom: PD186062; HCV\_NSI; 1.  
 KM Polypeptide: Glycoprotein: Coat protein; Envelope protein;  
 KM Transmembrane; Nonstructural protein.  
 FT NON\_TER 1 1

FT CHAIN <1 63 MATRIX PROTEIN (POTENTIAL).  
 FT CHAIN 64 255 MAJOR ENVELOPE PROTEIN E (POTENTIAL).  
 FT CHAIN 256 >309 NONSTRUCTURAL PROTEIN NS1/E2 (POTENTIAL).  
 FT CARBOHYD 68 68 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 81 81 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 106 106 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 177 177 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 289 289 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 295 295 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 302 302 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT NON\_TER 309 309 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 SQ SEQUENCE 309 AA; 32922 MW; 6E858E9C3D0B9EA9 CRC64;  
 Alignment Scores:  
 Pred. No.: 5.05e-76 Length: 309  
 Score: 928.00 Matches: 165  
 Percent Similarity: 91.41% Conservative: 16  
 Best Local Similarity: 83.33% Mismatches: 17  
 Query Match: 80.35% Indels: 0  
 DB: 1 Gaps: 0  
 US-09-899-303a-7 (1-633) x POLG\_HCVH7 (1-309)  
 QY 34 GCGTTCGCGGACGTCATGAGGAGTACATTCGCTGCGGCGGCGGCGGCGGCGGCGGCGG 93  
 Db 1 GlyPheAlaAspLeuMetGlyTyrIleProLeuValGlyAlaProLeuGlyGlyAlaAla 20  
 QY 94 AGAGCCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 153  
 Db 21 ArgAlaLeuAlaHisGlyValArgValIleuGlnAspGlyValAlaThrAlaThrHis 40  
 QY 154 TTGCGGTGCTGCTTCTCTATCA 213  
 Db 41 LeuProGlyCysSerPheSerIlePheLeuLeuAlaLeuLeuSerCysLeuThrValPro 60  
 QY 214 GCTTCCCGCTATGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 273  
 Db 61 AlaSerAlaTyrGlnValAlaArgAsnSerSerGlyTyrHisValThrAsnAspCysPro 80  
 QY 274 AACCAAGCATTTGATGAGAGCGGAGCATGATGACGACGACGACGACGACGACGACGACGAC 333  
 Db 81 AsnSerSerIleValTyrGlnThrAlaAspThrIleLeuHisSerProGlyCysValPro 100  
 QY 334 TGCCTTGGGAGAACACTCTTCCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 393  
 Db 101 CysValAlaArgGlnGlyAsnAlaSerIleCysTyrValProValAlaProThrValAlaThr 120  
 QY 394 AGGAACGCGACGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 453  
 Db 121 ArgAspAlaAsnLeuProAlaThrGlnLeuAlaArgHisIleAspLeuValGlySer 140  
 QY 454 GCTCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 513  
 Db 141 AlaThrLeuCysSerAlaLeuTyrValGlyAspLeuGlySerValPheLeuValGly 160  
 QY 514 CAGCTTTCACCATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 573  
 Db 161 GlnLeuPheThrPheSerProArgHisThrThrGlnAspCysAsnCysSerIle 180  
 QY 574 TATCCCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 627  
 Db 181 TyrProGlyHisIleThrGlnHisArgMetAlaTrpAspMetMetLeuTyr 198  
 RESULT 13  
 POLG\_HCVH1  
 ID POLG\_HCVH1 STANDARD: PRT: 192 AA.  
 AC P27954;  
 DT 01-AUG-1992 (Rel. 23, Created)  
 DT 01-AUG-1992 (Rel. 23, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Genome polypeptide [Contains: Matrix protein (Envelope protein M);  
 DE Major envelope protein E] (Fragment).  
 OS Hepatitis C virus (isolate EC1) (HCV).







GenCore version 5.1.3  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 18, 2003, 05:54:54 ; Search time 46 Seconds  
(without alignments)  
4220.139 Million cell updates/sec

Title: US-09-899-303A-7

Perfect score: 633  
Sequence: 1 ATGTTGGGTAAAGTCAATCGA.....TGATGATGACTGTAATAG 633

Scoring table: IDENTITY\_NUC  
Gapop 10.0, Gapext 1.0

Searched: 441362 seqs, 153338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued\_Patents\_NA.\*  
1: /cgn2\_6/ptodata/1/lna/5A.COMB.seq:\*  
2: /cgn2\_6/ptodata/1/lna/5B.COMB.seq:\*  
3: /cgn2\_6/ptodata/1/lna/6A.COMB.seq:\*  
4: /cgn2\_6/ptodata/1/lna/6B.COMB.seq:\*  
5: /cgn2\_6/ptodata/1/lna/PTUS.COMB.seq:\*  
6: /cgn2\_6/ptodata/1/lna/backfiles1.seq:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	633	100.0	633	3	US-08-612-973-7
2	633	100.0	633	4	US-08-927-597-7
3	612.8	96.8	636	3	US-08-612-973-13
4	612.8	96.8	636	4	US-08-927-597-13
5	608.8	96.2	795	3	US-08-612-973-5
6	608.8	96.2	795	4	US-08-927-597-5
7	605.8	95.7	2082	4	US-08-612-973-47
8	605.8	95.7	2082	4	US-08-927-597-47
9	605.8	95.7	2433	3	US-08-612-973-49
10	605.8	95.7	2433	4	US-08-927-597-49
11	561.4	88.7	932	1	US-08-081-072-15
12	561.4	88.7	932	1	US-08-449-093A-15
13	561.4	88.7	1539	2	US-08-470-426B-17
14	561.4	88.7	1863	2	US-08-470-426B-14
15	556.2	87.9	1037	1	US-08-462-195-1
16	556.2	87.9	1037	2	US-08-636-883-1
17	556.2	87.9	1037	3	US-08-127-829-1
18	553.4	87.4	2116	4	US-08-191-160-21
19	553.4	87.4	9599	3	US-09-014-416-4
20	553.4	87.4	9599	3	US-09-014-416-6
21	535.4	84.6	1167	1	US-08-324-977-9
22	535.4	84.6	1167	2	US-08-384-616-9
23	535.4	84.6	1167	2	US-08-904-686A-9
24	535.4	84.6	1167	2	US-09-315-850-9
25	535.4	84.6	1499	1	US-08-324-977-3
26	535.4	84.6	1499	2	US-08-384-616-3
27	535.4	84.6	1499	2	US-08-904-686A-3

28	535.4	84.6	1499	4	US-09-315-850-3	Sequence 3, Appl1
29	535.4	84.6	6039	1	US-08-324-977-11	Sequence 11, Appl1
30	535.4	84.6	6039	2	US-08-384-616-11	Sequence 11, Appl1
31	535.4	84.6	6039	2	US-08-904-686A-11	Sequence 11, Appl1
32	535.4	84.6	6039	4	US-09-315-850-11	Sequence 11, Appl1
33	535.4	84.6	9030	1	US-08-324-977-13	Sequence 13, Appl1
34	535.4	84.6	9030	2	US-08-384-616-13	Sequence 13, Appl1
35	535.4	84.6	9030	2	US-08-904-686A-13	Sequence 13, Appl1
36	535.4	84.6	9030	4	US-09-315-850-13	Sequence 13, Appl1
37	535.4	84.6	9416	1	US-08-324-977-1	Sequence 1, Appl1
38	535.4	84.6	9416	2	US-08-384-616-1	Sequence 1, Appl1
39	535.4	84.6	9416	2	US-08-904-686A-1	Sequence 1, Appl1
40	535.4	84.6	9416	4	US-09-315-850-1	Sequence 1, Appl1
41	535.4	84.6	9416	4	US-08-823-895A-27	Sequence 27, Appl1
42	534.2	84.4	1037	1	US-08-462-195-3	Sequence 3, Appl1
43	534.2	84.4	1037	2	US-08-636-883-3	Sequence 3, Appl1
44	534.2	84.4	1037	3	US-09-127-829-3	Sequence 3, Appl1
45	506.6	80.0	742	1	US-08-081-072-18	Sequence 18, Appl1

## ALIGNMENTS

RESULT 1  
US-08-612-973-7  
; Sequence 7, Application US/08612973  
; Patent No. 6150134  
GENERAL INFORMATION:  
APPLICANT: MAERTENS, GEERT  
APPLICANT: BOSMAN, FONS  
APPLICANT: DE MARTYNOFF, GUY  
APPLICANT: BUYSE, MARIE-ANGE  
TITLE OF INVENTION: PURIFIED HEPATITIS C VIRUS ENVELOPE  
TITLE OF INVENTION: PROTEINS FOR DIAGNOSTIC AND THERAPEUTIC USE  
NUMBER OF SEQUENCES: 111  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: NIXON & VANDERHUYE P.C.  
STREET: 1100 NORTH GLEBE ROAD  
CITY: ARLINGTON  
STATE: VIRGINIA  
COUNTRY: U.S.A.  
ZIP: 22201-4714  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25 (EPO)  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/612,973  
FILING DATE: 11-MAR-1996  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: BYRNE, THOMAS E.  
REGISTRATION NUMBER: 32,205  
REFERENCE/DOCKET NUMBER: 1487-10  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (703) 816-4000  
TELEFAX: (703) 816-4100  
INFORMATION FOR SEQ ID NO: 7:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 633 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: CDNA  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 1..630  
FEATURE:  
NAME/KEY: mat\_peptide  
LOCATION: 1..627

US-08-612-973-7

Query Match 100.0%; Score 633; DB 3; Length 633;  
Best Local Similarity 100.0%; Pred. No. 1.9e-164;  
Matches 633; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
QY 1 ATGTGGGTAAAGTCATGATACCTTACGTGCGCTTCCGACCTCATGGGGTACATT 60
    |||||||
Db 1 ATGTGGGTAAAGTCATGATACCTTACGTGCGCTTCCGACCTCATGGGGTACATT 60
QY 61 CCGCTCGTGGGGCCCCCTAGGGGGTCTGCGACAGCCCTGGGGCATGGGGTCCGGGTT 120
    |||||||
Db 61 CCGCTCGTGGGGCCCCCTAGGGGGTCTGCGACAGCCCTGGGGCATGGGGTCCGGGTT 120
QY 121 CTGGAAGACGGCGTGAATATGCAACAGGAAATTGCTGTTGCTTTCTATCTTC 180
    |||||||
Db 121 CTGGAAGACGGCGTGAATATGCAACAGGAAATTGCTGTTGCTTTCTATCTTC 180
QY 181 CTCTTGGCTTACTCTCTGCTGTGACCATTCGAGCTTCCGCTTATGAGGTGGCAGCTG 240
    |||||||
Db 181 CTCTTGGCTTACTCTCTGCTGTGACCATTCGAGCTTCCGCTTATGAGGTGGCAGCTG 240
QY 241 TCCGGGATGATCATGTATGACAGACATGCTCCACTCAGCATGTTGATGAGGACAGG 300
    |||||||
Db 241 TCCGGGATGATCATGTATGACAGACATGCTCCACTCAGCATGTTGATGAGGACAGG 300
QY 301 GACATGATCATGACACACCCCGGGTGCCTGCTGCTGGGAGAACACTCTTCCGC 360
    |||||||
Db 301 GACATGATCATGACACACCCCGGGTGCCTGCTGCTGGGAGAACACTCTTCCGC 360
QY 361 TGCTGGGTAGCGCTACCCCGACGCTGCGAGTAGGAAGCCGAGCTCCCACTACGACA 420
    |||||||
Db 361 TGCTGGGTAGCGCTACCCCGACGCTGCGAGTAGGAAGCCGAGCTCCCACTACGACA 420
QY 421 ATACAGCCCAAGTGTGCTGCTGCTGGGGGCTGCTTCTGCTTCCGCTATGACG 480
    |||||||
Db 421 ATACAGCCCAAGTGTGCTGCTGCTGGGGGCTGCTTCTGCTTCCGCTATGACG 480
QY 481 GGGGATCTCTGGGATCTGTCTCTCTGCTGCCAGCTGTCCACATCTCGGCTCCGCG 540
    |||||||
Db 481 GGGGATCTCTGGGATCTGTCTCTCTGCTGCCAGCTGTCCACATCTCGGCTCCGCG 540
QY 541 CATGAGACGGGTGACAGATGCAATTGCTCAATCTATCCCGGCACATTAACAGTACCGT 600
    |||||||
Db 541 CATGAGACGGGTGACAGATGCAATTGCTCAATCTATCCCGGCACATTAACAGTACCGT 600
QY 601 ATGGCTGGGATATGATGATGAATGTAATAG 633
    |||||||
Db 601 ATGGCTGGGATATGATGATGAATGTAATAG 633
```

RESULT 2  
US-08-927-597-7

Sequence 7, Application US/08927597

Patent No. 6,245,503

GENERAL INFORMATION:

APPLICANT: MAERTENS, GEERT

APPLICANT: BOSMAN, FONS

APPLICANT: DE MARTYNOFF, GUY

APPLICANT: BUYSSE, MARIE-ANGE

TITLE OF INVENTION: PURIFIED HEPATITIS C VIRUS ENVELOPE

NUMBER OF SEQUENCES: 111

CORRESPONDENCE ADDRESS: 111

ADDRESS: NIXON &amp; VANDERHAYE P.C.

STREET: 1100 NORTH GLEBE ROAD

CITY: ARLINGTON

STATE: VIRGINIA

COUNTRY: U.S.A.

ZIP: 22201-4714

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25 (BPO)  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/927,597  
FILING DATE:

CLASSIFICATION:

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/612,973

FILING DATE: 11-MAR-1996

ATTORNEY/AGENT INFORMATION:

NAME: BYRNE, THOMAS E.

REGISTRATION NUMBER: 32,205

REFERENCE/DOCKET NUMBER: 1487-10

TELECOMMUNICATION INFORMATION:

TELEPHONE: (703) 816-4100

TELEFAX: (703) 816-4100

INFORMATION FOR SEQ ID NO: 7:

SEQUENCE CHARACTERISTICS:

LENGTH: 633 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

MOLECULE TYPE: cDNA

TOPOLOGY: linear

HYPOTHETICAL: NO

ANTI-SENSE: NO

FEATURE:

NAME/KEY: CDS

LOCATION: 1..630

FEATURE:

NAME/KEY: mat\_peptide

LOCATION: 1..627

US-08-927-597-7

Query Match 100.0%; Score 633; DB 4; Length 633;  
Best Local Similarity 100.0%; Pred. No. 1.9e-164;  
Matches 633; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
QY 1 ATGTGGGTAAAGTCATGATACCTTACGTGCGGCTTCCGACCTCATGGGGTACATT 60
    |||||||
Db 1 ATGTGGGTAAAGTCATGATACCTTACGTGCGGCTTCCGACCTCATGGGGTACATT 60
QY 61 CCGCTCGTGGGGCCCCCTAGGGGGTCTGCGACAGCCCTGGGGCATGGGGTCCGGGTT 120
    |||||||
Db 61 CCGCTCGTGGGGCCCCCTAGGGGGTCTGCGACAGCCCTGGGGCATGGGGTCCGGGTT 120
QY 121 CTGGAAGACGGCGTGAATATGCAACAGGAAATTGCTGTTGCTTTCTATCTTC 180
    |||||||
Db 121 CTGGAAGACGGCGTGAATATGCAACAGGAAATTGCTGTTGCTTTCTATCTTC 180
QY 181 CTCTTGGCTTACTCTCTGCTGTGACCATTCGAGCTTCCGCTTATGAGGTGGCAGCTG 240
    |||||||
Db 181 CTCTTGGCTTACTCTCTGCTGTGACCATTCGAGCTTCCGCTTATGAGGTGGCAGCTG 240
QY 241 TCCGGGATGATCATGTATGACAGACATGCTCCACTCAGCATGTTGATGAGGACAGG 300
    |||||||
Db 241 TCCGGGATGATCATGTATGACAGACATGCTCCACTCAGCATGTTGATGAGGACAGG 300
QY 301 GACATGATCATGACACACCCCGGGTGCCTGCTGCTGGGAGAACACTCTTCCGC 360
    |||||||
Db 301 GACATGATCATGACACACCCCGGGTGCCTGCTGCTGGGAGAACACTCTTCCGC 360
QY 361 TGCTGGGTAGCGCTACCCCGACGCTGCGAGCTAGGAAGCCGACAGCTCCCACTACGACA 420
    |||||||
Db 361 TGCTGGGTAGCGCTACCCCGACGCTGCGAGCTAGGAAGCCGACAGCTCCCACTACGACA 420
QY 421 ATACAGCCCAAGTGTGCTGCTGCTGGGGGCTGCTTCTGCTTCCGCTATGACG 480
    |||||||
Db 421 ATACAGCCCAAGTGTGCTGCTGCTGGGGGCTGCTTCTGCTTCCGCTATGACG 480
QY 481 GGGGATCTCTGGGATCTGTCTCTCTGCTGCCAGCTGTCCAGCTGTCACCATCTCGCGCGG 540
    |||||||
Db 481 GGGGATCTCTGGGATCTGTCTCTCTGCTGCCAGCTGTTCACCATCTCGCGCGG 540
```





Db 301 GACATGATCATGCACACCCCGGGTGCCTGCCCTGCTCGGAGAGAACACTCTTCCCGC 360  
 Oy 361 TGTGGGTAGGCTACACCCCGACCTGCGACCTAGAAAGCCAGCTCCCACTACAGACA 420  
 Db 361 TGTGGGTAGGCTACACCCCGACCTGCGACCTAGAAAGCCAGCTCCCACTACAGACA 420  
 Oy 421 ATACGAGCCACGTCGATTTCTGCTGGGGCGGCTGCTTTCTGTTCCGCTATGACGTG 480  
 Db 421 ATACGAGCCACGTCGATTTCTGCTGGGGCGGCTGCTTTCTGTTCCGCTATGACGTG 480  
 Oy 481 GGGGATCTCTCGGATCTGCTCTGCTGCTCCAGCTGTCACACATCGCCCTGCGCGG 540  
 Db 481 GGGGATCTCTCGGATCTGCTCTGCTGCTCCAGCTGTCACACATCGCCCTGCGCGG 540  
 Oy 541 CATGAGAGGTCGAGAGCTGCAATTCATCTATCCGCGCACATTAACAGTACCGT 600  
 Db 541 CATGAGAGGTCGAGAGCTGCAATTCATCTATCCGCGCACATTAACAGGTCACCGT 600  
 Oy 601 ATGCTTGGCATATGATGATGAATGCTGT 628  
 Db 601 ATGCTTGGCATATGATGATGAATGCTGT 628

# RESULT 6

; Sequence 5, Application US/08927597  
 ; Patent No. 6245503  
 ; GENERAL INFORMATION:  
 ; APPLICANT: MAERTENS, GEERT  
 ; APPLICANT: BOSMAN, FONS  
 ; APPLICANT: DE MARTYNOFF, GUY  
 ; APPLICANT: BUYSSE, MARIE-ANGE  
 ; TITLE OF INVENTION: PURIFIED HEPATITIS C VIRUS ENVELOPE  
 ; NUMBER OF SEQUENCES: 111  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: NIXON & VANDERHAYE P.C.  
 ; STREET: 1100 NORTH GLEBE ROAD  
 ; CITY: ARLINGTON  
 ; STATE: VIRGINIA  
 ; COUNTRY: U.S.A.  
 ; ZIP: 22201-4714  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: Patent Release #1.0, Version #1.25 (EPO)  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/08/927,597  
 ; FILING DATE:  
 ; CLASSIFICATION:  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: US 08/612,973  
 ; FILING DATE: 11-MAR-1996  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: BYRNE, THOMAS E.  
 ; REGISTRATION NUMBER: 32,205  
 ; REFERENCE/DOCKET NUMBER: 1487-10  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: (703) 816-4000  
 ; TELEFAX: (703) 816-4100  
 ; INFORMATION FOR SEQ ID NO: 5:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 795 base pairs  
 ; TYPE: nucleic acid  
 ; STRANDEDNESS: single  
 ; TOPOLOGY: linear  
 ; MOLECULE TYPE: cDNA  
 ; HYPOTHEetical: NO  
 ; ANTI-SENSE: NO  
 ; FEATURE:  
 ; NAME/KEY: CDS  
 ; LOCATION: 1..792

; FEATURE:  
 ; NAME/KEY: mat\_peptide  
 ; LOCATION: 1..789  
 ; US-08-927-597-5

Query Match 96.28; Score 608.8; DB 4; Length 795;  
 Best Local Similarity 98.18; Pred. No. 8.6e-158;  
 Matches 616; Conservative 0; Mismatches 12; Indels 0; Gaps 0;

Oy 1 ATGTTGGGTAGGTCATGATACCTTACGTCGGGCTTCCGCGACCTTCATGGGTACATT 60  
 Db 1 ATGTTGGGTAGGTCATGATACCTTACGTCGGGCTTCCGCGACCTTCATGGGTACATT 60  
 Oy 61 CCGCTCTGTCGGGCGCCCGCTAGGGGGTCTGTCAGAGCCCTCGGCGCATGCGGCTT 120  
 Db 61 CCGCTCTGTCGGGCGCCCGCTAGGGGGTCTGTCAGAGCCCTCGGCGCATGCGGCTT 120  
 Oy 121 CTGGAAGACGGCTGTAATATGCAACAGGAATTTGCTGTTCTCTTCTATCTTC 180  
 Db 121 CTGGAAGACGGCTGTAATATGCAACAGGAATTTGCTGTTCTCTTCTATCTTC 180  
 Oy 181 CTCTGGCTTACTGCTCTGCTGCTGACCAATCCAGCTTCCGCTATAGGTGCGAAGTG 240  
 Db 181 CTCTGGCTTACTGCTCTGCTGCTGACCAATCCAGCTTCCGCTATAGGTGCGAAGTG 240  
 Oy 241 TCCGGATGTACATGTCACAGACGACTGCTCAACTCAACTCAACTGTTATAGAGCAGC 300  
 Db 241 TCCGGATGTACATGTCACAGACGACTGCTCAACTCAACTCAACTGTTATAGAGCAGC 300  
 Oy 301 GACATGATCATGCACACCCCGGGTGCCTGCTGCTGCTGCGGAGAACACTCTTCCCGC 360  
 Db 301 GACATGATCATGCACACCCCGGGTGCCTGCTGCTGCTGCGGAGAACACTCTTCCCGC 360  
 Oy 361 TGTGGGTAGGCTACACCCCGACCTGCGACCTAGGAAGCCAGCGTCCCACTACAGACA 420  
 Db 361 TGTGGGTAGGCTACACCCCGACCTGCGACCTAGGAAGCCAGCGTCCCACTACAGACA 420  
 Oy 421 ATACGAGCCACGTCGATTTCTGCTGGGGCGGCTGCTTTCTGTTCCGCTATGACGTG 480  
 Db 421 ATACGAGCCACGTCGATTTCTGCTGGGGCGGCTGCTTTCTGTTCCGCTATGACGTG 480  
 Oy 481 GGGGATCTCTCGGATCTGCTCTGCTGCTCCAGCTGTCACACATCGCCCTGCGCGG 540  
 Db 481 GGGGATCTCTCGGATCTGCTCTGCTGCTCCAGCTGTCACACATCGCCCTGCGCGG 540  
 Oy 541 CATGAGAGGTCGAGAGCTGCAATTCATCTATCCGCGCACATTAACAGTACCGT 600  
 Db 541 CATGAGAGGTCGAGAGCTGCAATTCATCTATCCGCGCACATTAACAGGTCACCGT 600  
 Oy 601 ATGCTTGGCATATGATGATGAATGCTGT 628  
 Db 601 ATGCTTGGCATATGATGATGAATGCTGT 628

## RESULT 7

; Sequence 47, Application US/08612973  
 ; Patent No. 6150134  
 ; GENERAL INFORMATION:  
 ; APPLICANT: MAERTENS, GEERT  
 ; APPLICANT: BOSMAN, FONS  
 ; APPLICANT: DE MARTYNOFF, GUY  
 ; APPLICANT: BUYSSE, MARIE-ANGE  
 ; TITLE OF INVENTION: PURIFIED HEPATITIS C VIRUS ENVELOPE  
 ; NUMBER OF SEQUENCES: 111  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: NIXON & VANDERHAYE P.C.  
 ; STREET: 1100 NORTH GLEBE ROAD  
 ; CITY: ARLINGTON  
 ; STATE: VIRGINIA  
 ; COUNTRY: U.S.A.  
 ; ZIP: 22201-4714

```
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentln Release #1.0, Version #1.25 (EPO)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/612,973
FILING DATE: 11-MAR-1996
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: BYRNE, THOMAS E.
REGISTRATION NUMBER: 32,205
REFERENCE/DOCKET NUMBER: 1487-10
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 816-4000
TELEFAX: (703) 816-4100
INFORMATION FOR SEQ ID NO: 47:
SEQUENCE CHARACTERISTICS:
LENGTH: 2082 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
HYPOTHETICAL: NO
ANTI-SENSE: NO
FEATURE:
NAME/KEY: CDS
LOCATION: 1..2079
FEATURE:
NAME/KEY: mat_peptide
LOCATION: 1..2076
US-08-612-973-47
```

Query Match 95.7%; Score 605.8; DB 3; Length 2082;  
Best Local Similarity 98.1%; Pred. No. 7.7e-157;  
Matches 613; Conservative 0; Mismatches 12; Indels 0; Gaps 0;

```
QY 4 TTGGGTAAGGTCAATCGATTACCTTACCTGCGGCTTCGCCGACCTCATGAGGGTACATTCCG 63
DB 4 TTGGGTAAGGTCAATCGATTACCTTACCTGCGGCTTCGCCGACCTCATGAGGGTACATTCCG 63
QY 64 CTGTGCGGCCCCCCCCCTAGAGGGGGTGCCTGCCAGAACCCCTGGCGCATGGCGCTTCG 123
DB 64 CTGTGCGGCCCCCCCCCTAGAGGGGGTGCCTGCCAGAACCCCTGGCGCATGGCGCTTCG 123
QY 124 GAAGACGGCGTGAACATGACACAGGGAATTTGGCTGTGCTCTTCTCTATCTTCTC 183
DB 124 GAAGACGGCGTGAACATGACACAGGGAATTTGGCGGTGCTCTTCTCTATCTTCTC 183
QY 184 TTGGGCTTGTCTGCTGTGACCATTCAGCTTCGCTTATGAGGTGCCAACGTGTCC 243
DB 184 TTGGGCTTGTCTGCTGTGACCGTTCCAGCTTCGCTTATGAGGTGCCAACGTGTCC 243
QY 244 GCGATGTACCATGTCAAGACGACTGTCTCAACTCAAGCATTTGTATGAGCAGCGGAC 303
DB 244 GCGATGTACCATGTCAAGACGACTGTCTCAACTCAAGCATTTGTATGAGCAGCGGAC 303
QY 304 ATGATCATGACACACCCCGGGGTGGTGCCTTGGGGAAGAACACTTCCCGCTGC 363
DB 304 ATGATCATGACACACCCCGGGGTGGTGCCTTGGGGAAGAACACTTCCCGCTGC 363
QY 364 TTGGTAGCGCTCAACCCCGACGCTCGCAGTAGAAGAACCCAGCGTCCCACTAGCAATA 423
DB 364 TTGGTAGCGCTCAACCCCGACGCTCGCAGTAGAAGAACCCAGCGTCCCACTAGCAATA 423
QY 424 CGAGCCGACGTCGATTTGCTGTGGGGCGGCTTTCGTTCCGCTATGATGAGTGGG 483
DB 424 CGAGCCGACGTCGATTTGCTGTGGGGCGGCTTTCGTTCCGCTATGATGAGTGGG 483
QY 484 GACCTCTGCGGATCTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 543
DB 484 GACCTCTGCGGATCTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 543
```

```
QY 544 GAGACGGTGCAGACTGTCATTCCTCATCTATCCGCCGACCATTAACAGGTACCCGTATG 603
DB 544 GAGACGGTGCAGACTGTCATTCCTCATCTATCCGCCGACCATTAACAGGTACCCGTATG 603
QY 604 GCTTGGATATGATGATGAACTGCT 628
DB 604 GCTTGGATATGATGATGAACTGCT 628
```

## RESULT 8

```
US-08-927-597-47
Sequence 47, Application US/08927597
Patent No. 6245503
GENERAL INFORMATION:
APPLICANT: MAERTENS, GEERT
APPLICANT: BOSMAN, FONS
APPLICANT: DE MARTYNOFF, GUY
APPLICANT: BUYS, MARIE-ANGE
TITLE OF INVENTION: PURIFIED HEPATITIS C VIRUS ENVELOPE
NUMBER OF SEQUENCES: 111
CORRESPONDENCE ADDRESS:
ADDRESSEE: NIXON & VANDERHYTE P.C.
STREET: 1100 NORTH GLEBE ROAD
CITY: ARLINGTON
STATE: VIRGINIA
COUNTRY: U.S.A.
ZIP: 22201-4714
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentln Release #1.0, Version #1.25 (EPO)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/927,597
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/612,973
FILING DATE: 11-MAR-1996
ATTORNEY/AGENT INFORMATION:
NAME: BYRNE, THOMAS E.
REGISTRATION NUMBER: 32,205
REFERENCE/DOCKET NUMBER: 1487-10
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 816-4000
TELEFAX: (703) 816-4100
INFORMATION FOR SEQ ID NO: 47:
SEQUENCE CHARACTERISTICS:
LENGTH: 2082 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
HYPOTHETICAL: NO
ANTI-SENSE: NO
FEATURE:
NAME/KEY: CDS
LOCATION: 1..2079
FEATURE:
NAME/KEY: mat_peptide
LOCATION: 1..2076
US-08-927-597-47
```

Query Match 95.7%; Score 605.8; DB 4; Length 2082;  
Best Local Similarity 98.1%; Pred. No. 7.7e-157;  
Matches 613; Conservative 0; Mismatches 12; Indels 0; Gaps 0;

```
QY 4 TTGGGTAAGGTCAATCGATTACCTTACCTGCGGCTTCGCCGACCTCATGAGGGTACATTCCG 63
DB 4 TTGGGTAAGGTCAATCGATTACCTTACCTGCGGCTTCGCCGACCTCATGAGGGTACATTCCG 63
QY 64 CTGTGCGGCCCCCCCCCTAGAGGGGGTGCCTGCCAGAACCCCTGGCGCATGGCGCTTCG 123
```

Db	64	CTCGTCGGCGCCCGCCCTATGGGGGGGGCGTGGCCAGGGGCCCTGGGGGCAATGGCGTCCGGGTCTC	123
OY	124	GAGACGGCGCTGAACCTATGCAACAGGGCAATTGCTGTGGTTCCTTTCTCTATCTTCTC	183
Db	124	GAGGACGGCGCTGAACCTATGCAACAGGGCAATTGGCCGGTGTCTCTTCTCTATCTTCTC	183
OY	184	TTGGCTTACTGTCCTGTCGTACCAATTCAGCTTCGGCTATATAGGTGGGCAACGTGTCC	243
Db	184	TTGGCTTGTGCTCTGCTGTCTGACCGTTCCAGCTTCCCGCTTATGAAGTGGCCAAAGTGTCC	243
OY	244	GGGATGTACCATGTTCAGCAACGACATGCTCCAACTCAAGCAATTGTATGTAGGCAAGGGAC	303
Db	244	GGGATGTACCATGTTCAGCAACGACATGCTCCAACTCAAGCAATTGTATGTAGGCAAGGGAC	303
OY	304	ATGATCATGACACACCCCGGGTGGCTGTGCTGCCCTTGGGGAGAACAACTTTCCCGCTGC	363
Db	304	ATGATCATGACACACCCCGGGTGGGTGCCCTTGGGGAGAACAACTTTCCCGCTGC	363
OY	364	TGGGTAGCGCTCACCCCGACGCTCGCAGTGTAGAACCCAGACGCTCCCACTAGACAAATA	423
Db	364	TGGGTAGCGCTCACCCCGACGCTCGCAGTGTAGAACCCAGACGCTCCCACTAGACAAATA	423
OY	424	CGAGCCCACTGTCGATTTTGCTGTTGGGGCGGCTGCTTGTCTTCGGCTATGTACGTGGGG	483
Db	424	CGAGCCCACTGTCGATTTTGCTGTTGGGGCGGCTGCTTGTCTTCGGCTATGTACGTGGGG	483
OY	484	GATGCTCTACCGATGTGTCCTTCCTGCTCCACAGCTGTTCACATCTTCGGCTCGCCGGCAT	543
Db	484	GATGCTCTACCGATGTGTCCTTCCTGCTTCACAGCTGTTCACATCTTCGGCTCGCCGGCAT	543
OY	544	GAGACGGGTGCAGGACATGCAATTTGCTCAATCTATCCCGGACACATTAACAGGTCAACGTATG	603
Db	544	GAGACGGGTGCAGGACATGCAATTTGCTCAATCTATCCCGGACACATTAACAGGTCAACGTATG	603
OY	604	GCTTGGGATATGATGATGAACGTGT	628
Db	604	GCTTGGGATATGATGATGAACGTGT	628

RESULT 9  
 US-08-612-973-49  
 Sequence 49, Application US/08612973  
 Patent No. 6150134  
 GENERAL INFORMATION:  
 APPLICANT: MAERTENS, GEERT  
 APPLICANT: BOSMAN, FONS  
 APPLICANT: DE MARTYNOFF, GUY  
 APPLICANT: BOYSE, MARIE-ANNE  
 TITLE OF INVENTION: PURIFIED HEPATITIS C VIRUS ENVELOPE  
 TITLE OF INVENTION: PROTEINS FOR DIAGNOSTIC AND THERAPEUTIC USE  
 NUMBER OF SEQUENCES: 111  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: NIXON & VANDERHAYE P. C.  
 STREET: 1100 NORTH GLEBE ROAD  
 CITY: ARLINGTON  
 STATE: VIRGINIA  
 COUNTRY: U.S.A.  
 ZIP: 22201-4714  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: Patentin Release #1.0, Version #1.25 (EPO)  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/612,973  
 FILING DATE: 11-MAR-1996  
 CLASSIFICATION: 435  
 ATTORNEY/AGENT INFORMATION:  
 NAME: BYRNE, THOMAS F.  
 REGISTRATION NUMBER: 32,205  
 REFERENCE/DOCKET NUMBER: 1487-10  
 TELECOMMUNICATION INFORMATION:

```

1  TELEPHONE: (703) 816-4000
2  TELEFAX: (703) 816-4100
3  INFORMATION FOR SNO ID NO: 49
4  SEQUENCE CHARACTERISTICS:
5  LENGTH: 2433 base pairs
6  TYPE: nucleic acid
7  STRANDEDNESS: single
8  TOPOLOGY: linear
9  MOLECULE TYPE: cDNA
10  HYPOTHETICAL: NO
11  ANTI-SENSE: NO
12  FEATURE:
13  NAME/KEY: CDS
14  LOCATION: 1..2430
15  FEATURE:
16  NAME/KEY: mat_peptide
17  LOCATION: 1..2427
18  JS-08-612-973-49

```

Query Match	95.78;	Score 605.8;	DB 3;	Length 2433;
Best Local Similarity	98.18;	Pred. No. 8e-157;		
Matches 613;	Conservative 0;	Mismatches 12;	Indels 0;	Gaps 0;

QY	4	TTGGGTAAAGGTCATCGATACCCCTACGTGGGGCTTCGCGACCTCATGGGGATACATCCG	63
Db	355	TTGGGTAAAGGTCATCGATACCCCTACATGGGGCTTCGCGACCTCATGGGGATACATCCG	414
QY	64	CTCTCGGGCGCCCCCTAAGGGGGTGTCCGAGAGCCCTGGCGCATGCGCTCCGGTCTG	123
Db	415	CTCTCGGGCGCCCCCTAAGGGGGGTGTCCGAGAGCCCTGGGGCATGGCGTCCGGGTCTG	474
QY	124	GAACACGGCGGTGAACCTATGCAACAGGGGAATTTGGCTGGTGTCTTTCTCTATCTTCTC	183
Db	475	GAGAGCGCGGTGAACCTATGCAACAGGGGAATTTGGCGGTGTCTTTCTCTATCTTCTC	534
QY	184	TTGGCTTATCTGCTCTGTCTGTACATTCACAGCTCCGCTTATGAGGTGGCCAGCGTGTCC	243
Db	535	TTGGCTTGTGCTGTCTGTCTGTACAGCTTCACAGCTTCGCTTATGAGGTGGCCAGCGTGTCC	594
QY	244	GGGATGTACACATGTACAGAACGACGTGCTCCAACTGATCAAGACTTGTATGAGGACGGGAC	303
Db	595	GGGATGTACCATGTACAGAACGACGTGCTCCAACTGATGTATGAGGACGGGAC	654
QY	304	ATGATCATGCAACACCCCGGGTGGGTGCCCTGGCTTTGGGGAGAACAACTTTCGCCGTGC	363
Db	655	ATGATCATGCAACACCCCGGGTGGGTGCCCTGGCTTTGGGGAGAACAACTTTCGCCGTGC	714
QY	364	TGGGTAGCGCTCACACCCACAGCTCGCAGTAGAAGCCAGACGTCGCCCATACGACAAATA	423
Db	715	TGGGTAGCGCTCACACCCACAGCTCGCAGTAGAAGCCAGACGTCGCCCATACGACAAATA	774
QY	424	CGACGCCACGTGCATTTTGCTGTTGGGGCGGCTGCTTCTGTCCGTATGTATGATCGTGGG	483
Db	775	CGACGCCACGTGCATTTTGCTGTTGGGGCGGCTGCTTCTGTCCGTATGTATGATCGTGGG	834
QY	484	GATCTCTGCGGATCTGTCTTCTCTGTCCAGCTGTTCACATCTGCGCCCGCGGCAT	543
Db	835	GATCTCTGCGGATCTGTCTTCTCTGTCTCCAGCTGTTCACATCTGCGCCCGCGGCAT	894
QY	544	GAGACGGGTGACGAGCATGCAATTTGCTCATATCATCCCGGCCACATATACAGTTCACCGTATG	603
Db	895	GAGACGGGTGACGAGCATGCAATTTGCTCATATCATCCCGGCCACATATACAGTTCACCGTATG	954
QY	604	GCTTTGGGATATGATGATGAACCTGCT	628
Db	955	GCTTTGGGATATGATGATGAACCTGCT	979

RESULT 10  
US-08-927-597-49  
; Sequence 49, Application US/08927597  
; Patent No. 6245503  
; GENERAL INFORMATION:

APPLICANT: MARTENS, GEERT  
APPLICANT: BOSMAN, FONS  
APPLICANT: DE MARTYNOFF, GUY  
APPLICANT: BUYSSE, MARIE-ANGE  
TITLE OF INVENTION: PURIFIED HEPATITIS C VIRUS ENVELOPE  
TITLE OF INVENTION: PROTEINS FOR DIAGNOSTIC AND THERAPEUTIC USE  
NUMBER OF SEQUENCES: 111  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: NIXON & VANDERHAYE P.C.  
STREET: 1100 NORTH GLEBE ROAD  
CITY: ARLINGTON  
STATE: VIRGINIA  
COUNTRY: U.S.A.  
ZIP: 22201-4714  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.25 (EPO)  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/927,597  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/612,973  
FILING DATE: 11-MAR-1996  
ATTORNEY/AGENT INFORMATION:  
NAME: BYRNE, THOMAS E.  
REGISTRATION NUMBER: 32,205  
REFERENCE/DOCKET NUMBER: 1487-10  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (703) 816-4000  
TELEFAX: (703) 816-4100  
INFORMATION FOR SEQ ID NO: 49:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2433 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 1..2430  
FEATURE:  
NAME/KEY: mat\_peptide  
LOCATION: 1..2427  
US-08-927-597-49  
Query Match 95.7%; Score 605.8; DB 4; Length 2433;  
Best Local Similarity 98.1%; Pred. No. 8e-157;  
Matches 613; Conservative 0; Mismatches 12; Indels 0; Gaps 0;

QY 4 TTGGGTAAAGTCATGATACCTTACGTGGGGCTTCCGCCACCTCATAGGGGTAAATTCG 63  
DB 355 TTGGGTAAAGTCATGATACCTTACGTGGGGCTTCCGCCACCTCATAGGGGTAAATTCG 414  
QY 64 CTGCTGGCGCCCCCTTAAGGGGGTGTGCTGCAAGAGCCCTGGGGCATGGCGTCCGGGTTCTG 123  
DB 415 CTGCTGGCGCCCCCTTAAGGGGGTGTGCTGCAAGAGCCCTGGGGCATGGCGTCCGGGTTCTG 474  
QY 124 GAAGACGGCGTGAATGCAACAGGGAATTTGCTGGTGGCTTTCTATCTTCTC 183  
DB 475 GAGAGCGCGGTAATGCAACAGGGAATTTGCGCGGTTGCTTTCTATCTTCTC 534  
QY 184 TTGGCTTTACTGTCTGTCTGACCATTCACGCTTCCGCTTATGAGGTGGCGCAAGTGTCC 243  
DB 535 TTGGCTTTACTGTCTGTCTGACCATTCACGCTTCCGCTTATGAGGTGGCGCAAGTGTCC 594  
QY 244 GGGATGATACATGTCAGAGCAAGTGTCTCAACTCAAGCAAGTGTGTATGAGGAGCGGAC 303  
DB 595 GGGATGATACATGTCAGAGCAAGTGTCTCAACTCAAGCAAGTGTGTATGAGGAGCGGAC 654

QY 304 ATGATCATGACACACCCCGGGGTGGTCCCTGCGTTGCGGAGAACACTCTTCCGCTGC 363  
DB 655 ATGATCATGACACACCCCGGGGTGGTCCCTGCGTTGCGGAGAACACTCTTCCGCTGC 714  
QY 364 TGGGTAGCGCTACACCCCGAGCTGCGAGCTAGAGAACCCAGCGTCCCATAGACATA 423  
DB 715 TGGGTAGCGCTACACCCCGAGCTGCGAGCTAGAGAACCCAGCGTCCCATAGACATA 774  
QY 424 CGAGCCACGTCGATTTGCTGTGGGGGCGCTCTTCTGTTCCGCTATGATAGTGAGG 483  
DB 775 CGAGCCACGTCGATTTGCTGTGGGGGCGCTCTTCTGTTCCGCTATGATAGTGAGG 834  
QY 484 GATCTGCGGATCTGTTCTGTTCCGCTGCTCCAGCTGTTACCAATCTGCGTCCGCGCAT 543  
DB 835 GACCTGCGGATCTGTTCTGTTCCGCTGCTCCAGCTGTTACCAATCTGCGTCCGCGCAT 894  
QY 544 GAGACGCTGCGAGCTGCAATGCTCAATCTATCTCCGCGCCACATTAACAGTCAACGATG 603  
DB 895 GAGACGCTGCGAGCTGCAATGCTCAATCTATCTCCGCGCCACATTAACAGTCAACGATG 954  
QY 604 GCTTGGATATGATGATGAATGCT 628  
DB 955 GCTTGGATATGATGATGAATGCT 979  
RESULT 11  
US-08-081-072-15  
Sequence 15, Application US/08081072  
Patent No. 5641654  
GENERAL INFORMATION:  
APPLICANT: No. 5641654oru MAKI, Kenjiro YAMAGUCHI, Ayumi  
APPLICANT: TOYOSHIMA, and Michinori KOHARA  
TITLE OF INVENTION: NON-A NON-B HEPATITIS-SPECIFIC  
TITLE OF INVENTION: ANTIGEN AND ITS USE IN HEPATITIS  
NUMBER OF SEQUENCES: 24  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: DiLe, Bronsteijn, Roberts & Cushman  
STREET: 130 Water Street  
CITY: Boston  
STATE: Massachusetts  
COUNTRY: USA  
ZIP: 02109-4280  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk, 3.50inch, 1.4MB storage  
COMPUTER: Apple Macintosh  
OPERATING SYSTEM: Macintosh Classic  
SOFTWARE: Microsoft Word Version 4.0A  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/081,072  
FILING DATE: June 22, 1993  
CLASSIFICATION: 435  
INFORMATION FOR SEQ ID NO: 15:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 932 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA to genomic RNA  
US-08-081-072-15  
Query Match 88.7%; Score 561.4; DB 1; Length 932;  
Best Local Similarity 93.5%; Pred. No. 8.7e-145;  
Matches 586; Conservative 0; Mismatches 41; Indels 0; Gaps 0;

QY 4 TTGGGTAAAGTCATGATACCTTACGTGGGGCTTCCGCCACCTCATAGGGGTAAATTCG 63  
DB 9 TTGGGTAAAGTCATGATACCTTACGTGGGGCTTCCGCCACCTCATAGGGGTAAATTCG 68  
QY 64 CTGCTGGCGCCCCCTTAAGGGGGTGTGCTGCAAGAGCCCTGGGGCATGGCGTCCGGGTTCTG 123  
DB 69 CTGCTGGCGCCCCCTTAAGGGGGTGTGCTGCAAGAGCCCTGGGGCATGGCGTCCGGGTTCTG 128

OY		124	GAAACGGCCGGAACATATGCAACAAGGGAAATTGGCTGTGGTGCCTTTCTCATCTTC	183
Dd		129	GAGNCGCGGTGACTATATGCAACAGGGNAATTTGCCGGTGCTTCTCTATCTTCTC	188
OY		184	TTGGCTTTACTGTCCTGTCGTGACCATTCCAGCTTCCGCTTATGAGTGGCCAACTGTCC	243
Dd		189	TTGGCTTTTTCGTCTGTCCTGTTTGAGCATTCOCACCTTCGCGCTTATGAGTGGCCAACTATCC	248
OY		244	GGAATGTACATGTGCACGAACGAGCTGCTCCAATCAGCATTTGTGTATGAGCAGCGGAC	303
Dd		249	GGGATATATACCATGTGCAGAACGAGACTGCTCCAACCTCAAGTATTTGTGTATGAGCAGCGGAC	308
OY		304	ATGATTCATGCACACCCCCGGGTGGTGGCCCGGTTGCGGAGAACAACCTGTCCCCTGC	363
Dd		309	ATGATTCATGCATACCCCCGGGTGGTGGCTGCTGCGCTTGGGGAACAACCTCTCCGTTGC	368
OY		364	TGGGTAGCGCTCACCCCCACGCTTCGACATAGAAACCCAGCGTCCCACTAGACATA	423
Dd		369	TGGCAGACGCTCACCTCCACGTTTAGCGGCGAGAAACACAGCGTCCCACTAGACATA	428
OY		424	CGACGCCAAGTCATTTTGCTGCTTGGGGCGGCTCTTTCGTTCGGCTATGTACGTGGG	483
Dd		429	CGAGCGCATGTGCATTTGTGCTGTGGGGCGGCTGCTTTCGCTCCGCTATGTACGTGGG	488
OY		484	GATCTTCGCGGATCTGTCCTTCCTCCGCTCCAGCTGTTCACACATCTGGCCCGCGGAT	543
Dd		489	GATCTCTGTGGATCTGTCTTCCTCTGTTTCCAGCTGTTCATCTTCTACCTGTGGCAT	548
OY		544	GAGACGGTCAGGACTGCAATTTGCTCATCTATCCCGGACCATATACAGGTCAACCGTATG	603
Dd		549	GAGACGATACGAGACTGCAATGCTCATCTATCCCGGACCATATACAGGTCAACCGTATG	608
OY		604	GCTTGGATATGATGATGAACTGGTAA	630
Dd		609	GCTTGGATATGATGATGAACTGGTCA	635
 RESULT 12 US-08-449-093A-15 Sequence 15, Application US/08449093A Patent No. 5662906 GENERAL INFORMATION: APPLICANT: NO. 5662906oru MAKI, Kenjiro YAMAGUCHI, Aymuni APPLICANT: TOYOSHIMA, and Michiori KOHARA TITLE OF INVENTION: NON-A-NON-B HEPATITIS-SPECIFIC TITLE OF INVENTION: ANTIGEN AND ITS USE IN HEPATITIS NUMBER OF SEQUENCES: 24 CORRESPONDENCE ADDRESS: ADDRESSEE: DIKE, BronsteIn, Roberts & Cushman, LLP STREET: 130 Water Street CITY: Boston STATE: Massachusetts COUNTRY: USA ZIP: 02109-4280 COMPUTER READABLE FORM: MEDIUM TYPE: Diskette COMPUTER: IBM Compatible OPERATING SYSTEM: DOS SOFTWARE: FASTSEQ Version 1.5 CURRENT APPLICATION DATA: APPLICATION NUMBER: US/08/449,093A FILING DATE: May 24, 1995 CLASSIFICATION: 424 PRIOR APPLICATION DATA: APPLICATION NUMBER: 08/081,072 FILING DATE: June 22, 1993 CLASSIFICATION: 424 PRIOR APPLICATION DATA: APPLICATION NUMBER: 07/726,141 FILING DATE: July 8, 1991 CLASSIFICATION: 424				

```

? INFORMATION FOR SEQ ID NO: 15:
? SEQUENCE CHARACTERISTICS:
?     LENGTH: 932 base pairs
?     TYPE: nucleic acid
?     STRANDEDNESS: double
?     TOPOLOGY: linear
? MOLECULE TYPE: cDNA to genomic RNA
? US-08-449-093A-15

Query Match      88.7%; Score 561.4; DB 1; Length 932;
Best Local Similarity 93.5%; Pred.No. 8,7e-145;
Matches 586; Conservative 0; Mismatches 41; Indels 0; Gaps 0;

DY          4   TTGGGTAAGTCAATCGATACCCCTTTAGCTGCAGGCTTCGCCGACCTCATGAGGGGTACATTCCG 63
              |||||
Dy          9   TTGGGTAAAGTCAATCATACCTCCTACATCGGCTTGCCGACTCATGGGTACATTCCG 68
              |||||
QY          64   CTCGTCCGCCCCCCCTTAGGGGGTGCCTGCCAGAGCCCTGGCCATGAGCGTCCGGATTCTG 123
              |||||
Dy          69   CTTGTCCGCCCCCCCTTAGGGGGTGCCTGCCAGAGCCCTGGCCACAATGATGTCCGGATTCTG 128
              |||||
QY          124  GAAGACGCGGTGAACATGCAACAGGAAATTTGGTCTGTGCTCTTCTCTCATCTCTCTC 183
              |||||
Dy          129  GAGAGCGCGGTGAACATGCAACAGGAAATTTGGTCTGTGCTCTTCTCTCATCTCTCTC 188
              |||||
QY          184   TTGGCTTTACTGTCTCCTGTCTGACCAATTCACGCTCCGCTTATAGAGTCCGCAAGCTGTCC 243
              |||||
Dy          189   TTGGCTTTGTGTGTCGATGTTTGAACATCCAGTTCGGTTATGAGTTCGCAAGCTATCC 248
              |||||
QY          244   GGGATGTACCATGTCAACGAACGACTGTCCAACCTCAAGCATTGTATAGAGCACAGGAC 303
              |||||
Dy          249   GGGATTAATACCATGTCAACGAACGACTGTCCAACCTCAAGTATTGTATAGAGCACAGGAC 308
              |||||
QY          304   ATGATCATGTGCACACCCTCCGGGTGCGTGGCCCTGCTTCCGGAGAACAATCTTCCCGCTGC 363
              |||||
Dy          309   ATGATCATGTGCATACCCCCTGGGTGCTGTGCCCTCTGCTTCCGGAGAACAATCTTCCCGCTGC 368
              |||||
QY          364   TGGGTAGCGCTCACCCCAACGCTCGCAGGTAGAAACGCCACAGCTCCACATAGACATA 423
              |||||
Dy          369   TGGGTAGCGCTCACTCCCACTGTAAGGGGCAAGAACACACAGGCTCCACATAGACATA 428
              |||||
QY          424   CGACGCCAGCTGATTTGGTCGTTGGGGGGGGTGCCTTCTGTCTCCGTATAGCTGGGG 483
              |||||
Dy          429   CGAGGGCATGTCAATTTGGTCGTTGGGGGGGGTGCCTTCTGTCTCCGTATAGCTGGGG 488
              |||||
QY          484   GATCTCTGCAGATCTGTCTTCTCTGCTGCTCCAGACTGTTCACCANTTCGCTCGCGGCA 543
              |||||
Dy          489   GATCTCTGTGATCTGTCTTCTCTGCTGCTCCAGCTGTTCACATCTCGTCCGGAT 548
              |||||
QY          544   GAGACGCTGCAGACATGCAATTCGTCAATCTATCCCGGCACATTAACAGTACCGGATG 603
              |||||
Dy          549   GAGACATTAACAGACATGCAATTCGTCAATCTATCCCGGCACATTAACAGTACCGGATG 608
              |||||
QY          604   GCPTGGATATGATGATGAACCTGTAA 630
              |||||
Dy          609   GCPTGGATATGATGATGAACCTGTCA 635
              |||||

RESULT 13
US-08-470-426B-17
: Sequence 17, Application US/08470426B
: Patent No. 5856458
: GENERAL INFORMATION:
: APPLICANT: Okamoto, Hiroaki
: TITLE OF INVENTION: OLIGONUCLEOTIDE PRIMERS, AND THEIR
: TITLE OF INVENTION: APPLICATION FOR HIGH-FIDELITY DETECTION OF NON-A-, NON-B-
: NUMBER OF SEQUENCES: 33
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Beverage, Degrandt, Weillacher & Young,
: ADDRESS: L.L.P.
: STREET: 1850 M Street, N.W., Suite 800
```

CITY: Washington  
STATE: DC  
COUNTRY: USA  
ZIP: 20036  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
OPERATING SYSTEM: IBM PC compatible  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/470,426B  
FILING DATE: 06-JUN-1995  
CLASSIFICATION: 536  
PRIORITY APPLICATION DATA:  
APPLICATION NUMBER: JP 2-153402  
FILING DATE: 12-JUN-1990  
ATTORNEY/AGENT INFORMATION:  
NAME: Wellacher, Robert G.  
REGISTRATION NUMBER: 20,531  
REFERENCE/DOCKET NUMBER: 06/59-47083.1  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (202) 659-2811  
TELEFAX: (202) 659-1462  
INFORMATION FOR SEQ. ID NO: 17:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1539 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: unknown  
MOLECULE TYPE: DNA (genomic)  
US-08-470-426B-17

Query Match 88.7%; Score 561.4; DB 2; Length 1539;  
Best local Similarity 93.5%; Pred. No. 1e-144;  
Matches 586; Conservative 0; Mismatches 41; Indels 0; Gaps 0;

QY 4 TTGGGTAAGGTCAATCGATACCTTACGTCGGGCTTGGCCCACTCATAGGGGTACATTCG 63  
DB 355 TTGGGTAAGGTCAATCGATACCTTACGTCGGGCTTGGCCCACTCATAGGGGTACATTCG 414  
QY 64 CTGCTGGGCGCCCGCTAGGGGGTGCCTGCAGAGCCCTGGCGCATGGCGTCCGGGTTCTG 123  
DB 415 CTGCTGGGCGCCCGCTAGGGGGTGCCTGCAGAGCCCTGGCGCATGGCGTCCGGGTTCTG 474  
QY 124 GAAGACGGCGTGAATGACAGAGGAAATTTGCTGTTCTTCTATCTTCTC 183  
DB 475 GAGAGCGGCGTGAATGACAGAGGAAATTTGCTGTTCTTCTATCTTCTC 534  
QY 184 TTGGCTTACTGTCCTGTCGACCATTCAGCTTCGCTTATAGAGTGGCGAAGTGTCC 243  
DB 535 TTGGCTTACTGTCCTGTCGACCATTCAGCTTCGCTTATAGAGTGGCGAAGTGTCC 594  
QY 244 GGGATGACCATGACAGAGAGAGTGTCCAACTCAAGCATTTGATGAGGAGCGGAC 303  
DB 595 GGGATGACCATGACAGAGAGAGTGTCCAACTCAAGCATTTGATGAGGAGCGGAC 654  
QY 304 ATGATCATGACACACCCCGGGGTGCTGCTGCTGCTGCGGAGAGAACTTCCCGCTGC 363  
DB 655 ATGATCATGACACACCCCGGGGTGCTGCTGCTGCTGCGGAGAGAACTTCCCGCTGC 714  
QY 364 TTGGTACGCTGACCCCGACGCTGCGAGCTAGAGAGCCGACGCTCCCACTAGACATA 423  
DB 715 TTGGTACGCTGACCCCGACGCTGCGAGCTAGAGAGCCGACGCTCCCACTAGACATA 774  
QY 424 CGACGCGACGATGATTTGCTGTTGGGGGCGCTTCTGTTCCGCTATGATGAGTGGG 483  
DB 775 CGACGCGACGATGATTTGCTGTTGGGGGCGCTTCTGTTCCGCTATGATGAGTGGG 834  
QY 484 GATCTCTGCGGATCTGCTCTCTCTCTCCACGCTGTTCACTTCCGCTCCGCGCAT 543  
DB 835 GATCTCTGCGGATCTGCTCTCTCTCTCCACGCTGTTCACTTCCGCTCCGCGCAT 894  
QY 544 GAGACGGTGCAGAGCTGCATTTGCTCATCTATTCGCGGCCACATTAACAGGTACCGTATG 603

DB 895 GAGACAGTGCAGAGCTGCATCTATTCGCGGCCATTTATCAGTCAACCGATG 954  
QY 604 GCTTGGATATGATGATGAGTGA 630  
DB 955 GCTTGGATATGATGATGAGTGA 981

RESULT 14  
US-08-470-426B-14  
Sequence 14, Application US/08470426B  
Patent No. 5856458

GENERAL INFORMATION:  
APPLICANT: Okamoto, Hiroaki  
APPLICANT: Nakamura, Tetsuo  
TITLE OF INVENTION: OLIGONUCLEOTIDE PRIMERS, AND THEIR  
TITLE OF INVENTION: APPLICATION FOR HIGH-FIDELITY DETECTION OF NON-A, NON-B  
TITLE OF INVENTION: HEPATITIS VIRUS  
NUMBER OF SEQUENCES: 33  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Beveridge, Degrandi, Wellacher & Young,  
ADDRESSEE: L.L.P.  
STREET: 1850 M Street, N.W., Suite 800  
CITY: Washington  
STATE: DC  
COUNTRY: USA

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
OPERATING SYSTEM: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/470,426B  
FILING DATE: 06-JUN-1995  
CLASSIFICATION: 536

PRIORITY APPLICATION DATA:  
APPLICATION NUMBER: JP 2-153402  
FILING DATE: 12-JUN-1990  
ATTORNEY/AGENT INFORMATION:  
NAME: Wellacher, Robert G.  
REGISTRATION NUMBER: 20,531  
REFERENCE/DOCKET NUMBER: 06/59-47083.1  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (202) 659-2811  
TELEFAX: (202) 659-1462  
INFORMATION FOR SEQ. ID NO: 14:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1863 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: unknown  
MOLECULE TYPE: DNA (genomic)  
US-08-470-426B-14

Query Match 88.7%; Score 561.4; DB 2; Length 1863;  
Best local Similarity 93.5%; Pred. No. 1e-144;  
Matches 586; Conservative 0; Mismatches 41; Indels 0; Gaps 0;

QY 4 TTGGGTAAGGTCAATCGATACCTTACGTCGGGCTTGGCCCACTCATAGGGGTACATTCG 63  
DB 679 TTGGGTAAGGTCAATCGATACCTTACGTCGGGCTTGGCCCACTCATAGGGGTACATTCG 738  
QY 64 CTGCTGGGCGCCCGCTAGGGGGTGCCTGCAGAGCCCTGGCGCATGGCGTCCGGGTTCTG 123  
DB 739 CTGCTGGGCGCCCGCTAGGGGGTGCCTGCAGAGCCCTGGCGCATGGCGTCCGGGTTCTG 798  
QY 124 GAAGACGGCGTGAATGACAGAGGAAATTTGCTGTTCTTCTATCTTCTC 183  
DB 799 GAGAGCGGCGTGAATGACAGAGGAAATTTGCTGTTCTTCTATCTTCTC 858  
QY 184 TTGGCTTACTGTCCTGTCGACCATTCAGCTTCGCTTATAGAGTGGCGAAGTGTCC 243  
DB 858 TTGGCTTACTGTCCTGTCGACCATTCAGCTTCGCTTATAGAGTGGCGAAGTGTCC 243

Db 859 TTGGCTTTGCTGCTGTTGACCATCCAGCTTCCGCTTATGTAAGTCCCAAGTGTCC 918  
 Oy 244 GGGATGATACATGTCAGAACGATGCTCCACTGATGATGATGAGGAGCGAC 303  
 Db 919 GGGATATACATGTCAGAACGATGCTCCACTGATGATGATGAGGAGCGAC 978  
 Oy 304 ATGATATGTCAGAACGATGCTCCACTGATGATGATGAGGAGCGAC 363  
 Db 979 ATGATATGTCAGAACGATGCTCCACTGATGATGATGAGGAGCGAC 1038  
 Oy 364 TTGGTAGCGCTACCCCGACGCTGAGAGTGAAGCCGAGCTCCCACTAGACATA 423  
 Db 1039 TTGGTAGCGCTACCCCGACGCTGAGAGTGAAGCCGAGCTCCCACTAGACATA 1098  
 Oy 424 CGAGCGCAGCTGATTTGCTGCTTGGGCGGCTGCTTCTGCTCCGATGATGAGGAG 483  
 Db 1099 CGAGCGCAGCTGATTTGCTGCTTGGGCGGCTGCTTCTGCTCCGATGATGAGGAG 1158  
 Oy 484 GATCTGCGGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 543  
 Db 1159 GATCTGCGGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1218  
 Oy 544 GATCTGCGGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 603  
 Db 1219 GATCTGCGGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1278  
 Oy 604 GCTTGGATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 630  
 Db 1279 GCTTGGATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1305

RESULT 15

US-08-462-195-1  
 : Sequence 1, Application US/08462195  
 : Patent No. 5789544

GENERAL INFORMATION:

APPLICANT: MIYAMURA, TATSUO  
 APPLICANT: SAITO, IZUMU  
 APPLICANT: MATSUDA, YOSHIHARU  
 APPLICANT: HONDA, YOSHIKAZU  
 APPLICANT: SEKI, MAROKO  
 TITLE OF INVENTION: METHOD FOR PRODUCING ECTOPROTEIN OF  
 TITLE OF INVENTION: HEPATITIS C VIRUS  
 NUMBER OF SEQUENCES: 9  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: OHLON, SPIVAK, MCLELLAND, MAIER & NEUSTADT,  
 STREET: P.C.  
 CITY: 1755 S. Jefferson Davis Highway, Suite 400  
 STATE: Virginia  
 COUNTRY: U.S.A.  
 ZIP: 22202

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk  
 OPERATING SYSTEM: IBM PC compatible  
 SOFTWARE: Patent Release #1.0, Version #1.25  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/462,195  
 FILING DATE: 05-JUN-1995  
 CLASSIFICATION: 514  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US/08/446,303  
 FILING DATE: 22-MAY-1995  
 APPLICATION NUMBER: US 08/074,584  
 FILING DATE: 11-JUN-1993  
 APPLICATION NUMBER: JP 152487/1992  
 FILING DATE: 11-JUN-1992  
 ATTORNEY/AGENT INFORMATION:  
 NAME: OHLON, NO. 5789544man F.  
 REGISTRATION NUMBER: 24,618  
 REFERENCE/DOCKET NUMBER: 4169-003-0  
 TELECOMMUNICATION INFORMATION:

TELEPHONE: (703) 413-3000  
 TELEFAX: (703) 413-2220  
 TELEX: 248855 OPAT UR  
 INFORMATION FOR SEQ ID NO: 1:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 1037 base pairs  
 TYPE: nucleic acid  
 STRANDEDNESS: double  
 TOPOLOGY: linear  
 MOLECULE TYPE: DNA (genomic)  
 ANTI-SENSE: NO  
 ORIGINAL SOURCE:  
 ORGANISM: Hepatitis C virus  
 IMMEDIATE SOURCE:  
 CLONE: pUC010  
 FEATURE:  
 NAME/KEY: CDS  
 LOCATION: 17..1036  
 US-08-462-195-1

Query Match

Best Local Similarity 93.1%; Score 556.2; DB 1; Length 1037;  
 Matches 582; Conservative 0; Mismatches 43; Indels 0; Gaps 0;

Oy 4 TTGGTAGGCTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 63  
 Db 371 TTGGTAGGCTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 430  
 Oy 64 CTGCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 123  
 Db 431 CTGCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 490  
 Oy 124 GAAGACGGGCTGAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 183  
 Db 491 GAAGACGGGCTGAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 550  
 Oy 184 TTGGCTTACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 243  
 Db 551 TTGGCTTACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 610  
 Oy 244 GGGATATACATGTCAGAACGATGCTCCACTGATGATGATGATGATGATGATGATGATGATGAT 303  
 Db 611 GGGATATACATGTCAGAACGATGCTCCACTGATGATGATGATGATGATGATGATGATGATGAT 670  
 Oy 304 ATGATATGTCAGAACGATGCTCCACTGATGATGATGATGATGATGATGATGATGATGATGATGAT 363  
 Db 671 ATGATATGTCAGAACGATGCTCCACTGATGATGATGATGATGATGATGATGATGATGATGATGAT 730  
 Oy 364 TTGGTAGCGCTACCCCGACGCTGAGAGTGAAGCCGAGCTCCCACTAGACATA 423  
 Db 731 TTGGTAGCGCTACCCCGACGCTGAGAGTGAAGCCGAGCTCCCACTAGACATA 790  
 Oy 424 CGAGCGCAGCTGATTTGCTGCTTGGGCGGCTGCTTCTGCTCCGATGATGAGGAG 483  
 Db 791 CGAGCGCAGCTGATTTGCTGCTTGGGCGGCTGCTTCTGCTCCGATGATGAGGAG 850  
 Oy 484 GATCTGCGGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 543  
 Db 851 GATCTGCGGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 910  
 Oy 544 GATCTGCGGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 603  
 Db 911 GATCTGCGGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 970  
 Oy 604 GCTTGGATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 628  
 Db 971 GCTTGGATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 995

Search completed: February 18, 2003, 07:08:51  
 Job time : 53 secs

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.3  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - protein search, using frame\_plus\_n2p model

Run on: February 12, 2003, 08:55:00 ; Search time 25.5 Seconds  
(without alignments)  
4772.791 Million cell updates/sec

Title: US-09-899-303a-7  
Sequence: 1 ATGTTGGTGAAGCTCATCGA.....TGATGATGAAGCTGTAATAG 633

Scoring table:  
BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 566448

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Command line parameters:  
-MODEL-frame+\_n2p.model -DEV-xip  
-O/cgn2\_1/USPTO.spool/US09899303/runat.12022003.085449.28276/app\_query.fasta.1.775  
-DB-PIR.73 -OPMT-fastan -SUFFIX-n2p.rpr -MINMATCH=0.1 -LOOPEXT=0  
-UNITS-bits -START=1 -END=1 -MATRIX-Biosum62 -TRANS-human40.cdi -LIST=45  
-DOCALIGN=200 -THR\_SCORE-pct -THR\_MAX=100 -THR\_MIN=0 -ALIGN=15 -MODE-LOCAL  
-OUTFMT=ptc -NORM-ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000  
-USRR=US09899303.@CGN.1.1.48.@runat.12022003.085449.28276 -NCPU=6 -ICPU=3  
-NO\_XLPEX -NO\_MMAL -LARGEROVER -NEG\_SCORES=0 -WAIT -LONGLOG -DEV\_TIMEOUT=120  
-MAIN\_TIMEOUT=30 -THREDS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7  
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

- 1: PIR.73:\*
- 2: PIR.73:\*
- 3: PIR.73:\*
- 4: PIR.73:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1094	94.7	513	2	PC1284
2	1084	93.9	441	2	S12707
3	1083	93.8	787	2	PN0677
4	1078	93.3	3010	1	GNMW7M
5	1073	92.9	3010	1	GNMW7M
6	1073	92.9	3010	1	GNMW7M
7	1067	92.4	3010	1	S18030
8	1065	92.2	782	2	S18031
9	1063	92.0	782	2	S18031
10	1060	91.8	369	2	S18032
11	1060	91.8	369	2	S18032
12	1060	91.8	3010	1	GNMW7C
13	1057	91.5	520	2	GNMW7C
14	1045	90.5	550	2	JH0711

15	1044	90.4	523	2	JQ1926	polypeptide - hepa
16	1043	90.3	782	2	S19875	genome polypeptide
17	995	86.1	3011	1	GNMW7C	genome polypeptide
18	992	85.9	3011	1	GNMW7C	genome polypeptide
19	986	85.4	513	2	A44150	structural protein
20	982	85.0	3011	1	S40770	genome polypeptide
21	980	84.8	640	2	JQ1584	genome polypeptide
22	921	79.7	322	2	JN0265	genome polypeptide
23	891	77.1	411	2	PC2061	genome polypeptide
24	887	76.8	411	2	S41288	genome polypeptide
25	881	76.3	411	2	PC2060	genome polypeptide
26	847	73.3	415	2	PC4407	envelope protein -
27	843	73.0	3014	1	JC5620	genome polypeptide
28	826	71.5	874	2	JQ0883	genome polypeptide
29	817	70.7	3033	1	GNMW7B	genome polypeptide
30	802	69.4	874	2	JQ0881	genome polypeptide
31	802	69.4	3033	1	JQ1303	genome polypeptide
32	731	63.3	180	2	PC1303	genome polypeptide
33	723	62.6	180	2	PC1304	genome polypeptide
34	721	62.4	180	2	PC1305	genome polypeptide
35	716	62.0	315	2	PS0164	envelope glycoprot
36	699	60.5	315	2	PS0165	envelope glycoprot
37	688	59.6	315	2	PS0165	envelope glycoprot
38	671	58.1	177	2	S32745	genome polypeptide
39	667	57.7	177	2	S25123	genome polypeptide
40	654	56.6	177	2	S32743	genome polypeptide
41	651	56.4	177	2	S32749	genome polypeptide
42	641	55.5	177	2	S32746	genome polypeptide
43	639	55.3	177	2	S32741	genome polypeptide
44	447.5	38.7	193	2	S35630	genome polypeptide
45	365	31.6	189	2	S32740	polypeptide - hepa

ALIGNMENTS

RESULT 1  
PC1284  
genome polypeptide - hepatitis C virus (isolate HC-J4) (fragment)  
C:Species: hepatitis C virus  
C:Date: 30-Sep-1993 #sequence\_revision 30-Sep-1993 #text\_change 17-Nov-2000  
C:Accession: PC1284  
R:Okamoto, H.; Okada, S.; Sugiyama, Y.; Yokumoto, S.; Tanaka, T.; Yoshizawa, H.; Tsu  
Jpn. J. Exp. Med. 60, 167-177, 1990  
A:Title: The 5'-terminal sequence of the hepatitis C virus genome.  
A:Reference number: PC1284; MUID:9103116; PMID:2170712  
A:Accession: PC1284  
A:Molecule type: genomic RNA  
A:Residues: 1-513 <OK>  
A:Cross-references: GB:D00832; NID:g221513; PIDN:BA00706.1; PID:g221514  
C:Superfamily: hepatitis C virus genome polypeptide

Alignment Scores:

Pred. No.: 1.63e-92  
Score: 1094.00  
Percent Similarity: 99.52%  
Best Local Similarity: 97.60%  
Query Match: 94.72%  
DB: 2  
Length: 513  
Matches: 203  
Conservative: 4  
Mismatches: 1  
Indels: 0  
Gaps: 0

US-09-899-303a-7 (1-633) x PC1284 (1-513)  
QY 4 TTGGTGAAGCTCATGATACCTTACGTCGCTTGCAGCATGATGGGATATTCG 63  
Db 119 LeuValIysValIleAspThrLeuThcCysdIyheAlaAspLeuMetGlyTyrIlePro 138  
QY 64 CTCGTGGCGCCGCCCTAGGGGGTGTCCAGAGCCCTGGGCGCATGGCGTGGGGTCTG 123  
Db 139 LeuValIglAlaProLeuIglGlyAlaIAlaArgAlaLeuAlaHisGlyValIArgValIleu 158  
QY 124 GAACAGCGCGTGAACATGACAGAGGAAATTTGGCTGCTTCTTCTCATTTCTCTC 183  
Db 159 GluAspGlyAlaAsnTyrAlaThrGlyAsnLeuProGlyCysSerPheSerIlePheLeu 178

OY	184	TTGGCTTACTGTCGTCTGCACCAATTCACAGCTCCCGCTTAATGAGTGCGCACAAGTGTCC	243
Db	179	LeuAlaIeuSerCysIleuThrIleProAlaSerAlaTygGIuValAlaArgInValSer	198
OY	244	GGAATGACATGTCACAGAACGAGCTGCCAACTCAAGCATTTGTATGAGCACCGAC	303
Db	199	GlyIleuTrnHisValIThrAsnAparCysSerAsnSerSelleValTygIuAlaAlaAsp	218
OY	304	ATGATCATGCACACCCCCGGGTGCGTCCCTGGTTCGGGAGACAACACTTTCCCGTGC	363
Db	219	MetIleMetHisTrnProGlyCysValProCysValAlaGluAspAsnSerArgLys	238
OY	364	TGGGTACGGCACCACCCAGCGTCGACGACGAGGAACGCCAGGTCGCCACTACGACATA	423
Db	239	TrpAlaIleuThrProThrIleuAlaAlaAlaArgAsnAlaSerValProthrThrIle	258
OY	424	CGACGCCACGTCGAATTGGTCGATTGGGGCGCGCTTTCTGTCTCGCATATGACCTGGGG	483
Db	259	ArgAlaTrnHisValAspIleuLeuValIGluAlaAlaIlaPheCysSerAlaMetIleTyAlaGly	278
OY	484	GATCTCTGCGGATGTGCTCTTCCTCGTCTCCCAAGCTGTACCAATCTCGCCTGCGCGGAT	543
Db	279	AspIeuCysGLySerValPheIeuValSerGlnIleuPheThrPheSerProArGArgHis	298
OY	544	GAGACGGTCAGACACTCAATTCGTCAATATATCCGGGCACATACAGTAGCAACGTTATG	603
Db	299	GluThrValIGlnAparCysAsnCysSerIleIeTyProGlyHisIleuSerGlyHisArgMet	318
OY	604	GCATTGGATATGATGATGAACGTGG	627
Db	319	AlaTrpAspMetIleMetLanTrp	326

```

RESULT 2
S12707
genome polypeptid - hepatitis C virus (fragment)
N:Contains: core protein; envelope protein
C:Species: hepatitis C virus
C:Date: 30-Sep-1993 #sequence_revision 30-Sep-1993 #text_change 17-Nov-2000
C:Accession: S12707
R:Rtauchil, K.; Kudo, Y.; Boommar, S.; Watanabe, Y.; Katayama, T.; Choo, Q.L.; Kuo, G.;
Nucleic Acids Res. 18, 4526, 1990
A:Title: Nucleotide sequence of core and envelope genes of the hepatitis C virus genome
A:Reference number: S12707; MUID:90356432; PMID:2117749
A:Accession: S12707
A:Molecule type: genomic RNA
A:Residues: 1-441 <TAK>
A:Cross-References: EMBL:D00574; NID:g221656; PIDN:BAAD0452.1; PID:g221657
C:Superfamily: hepatitis C virus genome polypeptid
C:Keywords: polypeptid

```

Alignment Scores:	
Pred. NO.:	1,35e-91
Score:	1094.00
Percent Similarity:	98.57%
Best Local Similarity:	95.67%
Query Match:	93.85%
DB:	2
Length:	441
Matches:	199
Conservative:	6
Mismatches:	3
Indels:	0
Gaps:	0

US-09-899-303A-7 (1-633) x SI2707 (1-441)

OY	4	TTGGTAGAGTCATCGATACCCCTTTAGCTGGGGCTTGCGGACCTCATGGGGTACATTCG	63
Db	119	LeuGlyLysValIleAspThrLeuThrCysGlyPheAlaIAspLeuMetGlyTyrIlePro	138
OY	64	CTCGTCGGGGCCCCCTTAGGGGGGTGCTGCCAGAACCTTGGCGCATGGCGTCCGGGTCTG	123
Db	139	LeuValGlyAlaProLeuGlyGlyAlaIleAlaTgAlaLeuAlaIleAspGlyAlaIATGValLeu	158
OY	124	GAAGACGGCGTGAACCTATGCAACAGGGGAATTTGGCTGGTGCCTTTCTGTATCTTCCTC	183
Db	159	GluAspGlyAlaSerTyrAlaThrGlyAsnLeuProGlyCysSerPheSerIleLeuLeu	178
OY	184	TTGGGTATACAGTCTCTGTCTGACCATTCACAGCTCCGCTTATGAGTGGCCCAACGTTGCC	243

Db	179	LeuLaIaLeuSerCysLeuThrIleProLaSerAlaTyGluValArgAsnValSer	198
Oy	244	GGATGATGACATGTCACAGAACGAGTCGTCCAACTCAAGCATTTGTGTAAGAGGACAGCGAC	303
Db	199	GlyIleIeuThrValIleIhrAsnAspCysSerAsnSerSerIleValItyTeluLaIaIaAsp	218
Oy	304	ATGATCATGACACACCCCGGGGTCCGAGCCCTGGTGTCGGGAGAACACACTCTTCCGGTGC	363
Db	219	ValIleMetHisIaIaProGlyCysValProCysValArgGluAsnAsnSerSerIaArgCys	238
Oy	364	TGGGTAGCGCTCACCCCCACGCTCGACAGCTAGGAACGCCAGCGTCCCACTACGACAATA	423
Db	239	TrpValaIaLeuThrProThrIleuLaIaAlaArgAsnLaSerValProThrThrLeu	258
Oy	424	CGAGCCACACGTCATTTGTCGCTGGGGGCGCGCTTTGTCTTCGGCATAGTACGTGGGG	483
Db	259	ArgArgHisValaAspLeuIeuValGlyThrAlaAlaIaPheCysSerAlaMetItyValGly	278
Oy	484	GATCTCGCGGATGTGTCCTGCTCGTCCGTCGCACAGCTTGACATGTGGCTGGCGCGGAT	543
Db	279	AspLeuTySglYserValPheLeuIleSerGlnLeuPheThrPheSerTrpArgArgHis	298
Oy	544	GAGACGGTGCAGAGACTGTCATTCCTCAATCTATCCGGGACACATAAGACAGTACCCGTATG	603
Db	299	GluThrValGlnAspCysAsnCysSerIleTyIleProGlyHisValSerCglYHisArgMet	318
Oy	604	GCCTTGCGATGATGATGAACGTGG	627
Db	319	AlaTrpAspMetIeuMetIaAsnTrp	326

RESULT 3  
PNU0677  
hypothetical protein 787 - hepatitis C virus (fragment)  
C:Species: hepatitis C virus  
C:Date: 14-Jul-1994 #sequence\_revision 14-Jul-1994 #text\_change 17-Nov-2000  
C:Accession: PNU0677  
R:Cho, S.H.; Yoon, J.I.; Chang, J.E.; Ahn, B.M.; Lee, C.H.; Lee, Y.I.  
Biochem. Biophys. Res. Commun. 196, 780-788, 1993  
A:Title: Genomic typing of hepatitis C viruses from Korean patients: Implications of  
A:Reference number: PNU0677; MUID:94059104; PMID:8240354  
A:Accession: PNU0677  
A:Molecule type: mRNA  
A:Residues: 1-787 <CHO>  
A:Cross-references: GB:I20498; NID:g1381031; PIND:AB02608.1; PID:g1381032  
C:Superfamily: hepatitis C virus genome polypeptide  
C:Keywords: glycoprotein; nonstructural protein  
F:196, 209, 234, 250, 305, 325, 421, 427, 452, 536, 544, 560, 580, 627, 649/Binding site: carboxyhyd

Alignment Scores:	
Pred. No.:	1,7e-91
Score:	1083.00
Percent Similarity:	99.04%
Best Local Similarity:	95.67%
Query Match:	93.77%
DB:	2
Gaps:	0

US-09-899-303A-7 (1-633) x PN0677 (1-787)

QY 4 TTGGGTAAAGGTCATCGATACCCCTTAAGTCGGGGCTTCGGCCGACCTCATAGGGGTACATCCG 63

Db 119 LeuGtLyvValIleAspThLeuThnCysGtLyPheAlaAspLeuMetGtLyTrIlePro 138

QY 64 CTCGTCGGGGCCCCCTCTAGGGGGTGTGCACAGACCCCTGGCGGATGCGTCGCCGGTCTG 123

Db 139 LeuValIglYlaProLeuGtLyGtYlaIalaIarGhIalaenIahIstGtYalaArgValLeu 158

QY 124 GAAGAGCGCGTGAACATATGCACAACGGGAATTGGCTGTGTCTTTCTCATCTTC 183

Db 159 GluAspGtYalaAsnTYlaIaThGtLyAsnLeuProGtLySerPheSerIlePheLeu 178

QY 184 TTGGCTTACCTGTCCTGTCTGACCATTCACAGTCCGCTTATGAGGTGGCAAGTGTCC 243

Db 179 LeuAlaLeuLeuSerCysLeuThrIleProAlaSerAlaTyrGluValArgAsnValSer 198  
 QY 244 GGGATGACCATGTCACGAGAGACTGCTCCAACTCAAGCATTTGTATGAGCAGCGAC 303  
 Db 199 GtlyAlaTyrHisValThrAsnAspCysSerAsnSerSerIleValTyrGluAlaAlaAsp 218  
 QY 304 ATGATCATGCACACACCCCGGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 363  
 Db 219 MetIleMetHisThrProGlyCysValProCysValArgGluAsnAsnPhSerSerArgCys 238  
 QY 364 TGGGTACCGCGACCCCGACGCTTCAGTTCAGTTCAGTTCAGTTCAGTTCAGTTCAGTTCAG 423  
 Db 239 TrpAlaLeuLeuThrProThrLeuAlaAlaArgAsnSerSerIleProThrThrIle 258  
 QY 424 CGAGCCGACGTCATGATTTGCTTCGTGGGGCGGCTTCCTTTCCTTCCTTCCTTCCTTCCTTC 483  
 Db 259 ArgArgHisIleAspLeuLeuValGlyAlaAlaAlaPheCysSerSerAlaMetTyrValGly 278  
 QY 484 GATCTTCGCGATCTGCTTCCTTCCTTCCTTCCTTCCTTCCTTCCTTCCTTCCTTCCTTCCT 543  
 Db 279 AspLeuCysGlySerValPheLeuValSerGlnLeuPheThrPheSerProArgArgTyr 298  
 QY 544 GAGCGGTGACGACGACATTTGCTTCGTGGGGCGGCTTCCTTCCTTCCTTCCTTCCTTCCTTC 603  
 Db 299 GluThrValGlnAspCysAsnCysSerIleTyrProGlyHisValSerGlyHisArgMet 318  
 QY 604 GCTTGGATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 627  
 Db 319 AlaTrpAspMetMetMetMetMetMetMetMetMetMetMetMetMetMetMetMetMet 326

RESULT 4  
 GNMVTW  
 genome polypeptide - hepatitis C virus (strain Taiwan)  
 N:Contains: capsid protein C; envelope protein M; hepacivirin (EC 3.4.21.98) (nonstructu

C:Species: hepatitis C virus  
 A:Note: host Homo sapiens (man)  
 C:Date: 31-Dec-1992 #sequence\_revision 31-Dec-1992 #text\_change 19-Jan-2001  
 C:Accession: A40244  
 R:Chem. P.J.; Lin, M.H.; Tai, K.F.; Liu, P.C.; Lin, C.J.; Chen, D.S.  
 Virology 188, 102-113, 1992  
 A:Title: The Taiwanese hepatitis C virus genome: sequence determination and mapping the  
 A:Reference number: A40244; MUID:92250206; PMID:1314449  
 A:Accession: A40244  
 A:Molecule type: genomic RNA  
 A:Residues: 1-3010 <CH>  
 A:Cross-references: GB:M84754  
 C:Superfamily: hepatitis C virus genome polypeptide  
 C:Keywords: ATP; capsid protein C; envelope protein M; glycoprotein; hydrolase; nonstructura  
 F:1-115/Product: capsid protein C #status predicted <CPC>  
 F:116-191/Product: envelope protein M #status predicted <EPM>  
 F:192-389/Product: major envelope protein E #status predicted <NEE>  
 F:390-729/Product: nonstructural protein NS1 #status predicted <NS1>  
 F:1007-1615/Product: nonstructural protein NS2 #status predicted <NS2>  
 F:1230-1237/Product: hepacivirin #status predicted <NS3>  
 F:1312-1317/Product: nucleotide-binding motif A (P-loop)  
 F:1316-1319/Region: nucleotide-binding motif B  
 F:1316-1319/Region: DEXH motif  
 F:1616-1862/Product: nonstructural protein NS4 #status predicted <N4A>  
 F:1863-2013/Product: nonstructural protein NS4 #status predicted <N4B>  
 F:2014-3010/Product: nonstructural protein NS5 #status predicted <NS5>  
 F:196-209,233,234,250,305,325,417,423,430,448,532,540,556,576,623,645,1213,1255,2041,207

Alignment Scores:  
 Pred. No.: 5, 09e-91 Length: 3010  
 Score: 1078.00 Matches: 200  
 Percent Similarity: 97.60% Conservative: 3  
 Best Local Similarity: 96.15% Mismatches: 5  
 Query Match: 93.33% Indels: 0  
 DB: 1 Caps: 0

US-09-899-303a-7 (1-633) x GNMVTW (1-3010)

QY 4 TTGGTAAAGTCATGATACCTTACGTCGCGCTTCGCGGACGTCATGAGGATACATTCG 63  
 Db 119 LeuGlyValIleAspThrLeuThrCysGlyPheAlaAspLeuMetGlyTyrIlePro 138  
 QY 64 CTCGTGGCGGCGCCCTAGGGGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 123  
 Db 139 LeuValGlyAlaProLeuGlyGlyValAlaArgAlaLeuAlaHisGlyValAlaValLeu 158  
 QY 124 GAAGACGCGGTGACATATGCAACAGGAAATTTGCTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 183  
 Db 159 GluAspGlyAlaAsnTyrAlaThrClyAsnLeuProGlyCysSerPheSerIlePheLeu 178  
 QY 184 TTGGCTTACTGTCCTGCTGACCATTCACGCTTCGCTTCCTTCCTTCCTTCCTTCCTTCCTTC 243  
 Db 179 LeuAlaLeuLeuSerCysLeuThrIleProAlaSerAlaTyrGluValHisAsnValSer 198  
 QY 244 GGGATGACCATGTCACGAGAGACTGCTCCAACTCAAGCATTTGTATGAGCAGCGAC 303  
 Db 199 GtlyAlaTyrHisValThrAsnAspCysSerAsnSerSerIleValTyrGluAlaAlaAsp 218  
 QY 304 ATGATCATGCACACACCCCGGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 363  
 Db 219 MetIleMetHisThrProGlyCysValProCysValArgGluAsnAsnPhSerSerArgCys 238  
 QY 364 TGGGTACCGCGACCCCGACGCTTCAGTTCAGTTCAGTTCAGTTCAGTTCAGTTCAGTTCAG 423  
 Db 239 TrpAlaLeuLeuThrProThrLeuAlaAlaArgAsnSerSerIleProThrThrIle 258  
 QY 424 CGAGCCGACGTCATGATTTGCTTCGTGGGGCGGCTTCCTTTCCTTCCTTCCTTCCTTCCTTC 483  
 Db 259 ArgArgHisIleAspLeuLeuValGlyAlaAlaAlaPheCysSerSerAlaMetTyrValGly 278  
 QY 484 GATCTTCGCGATCTGCTTCCTTCCTTCCTTCCTTCCTTCCTTCCTTCCTTCCTTCCTTCCT 543  
 Db 279 AspLeuCysGlySerValPheLeuValSerGlnLeuPheThrPheSerProArgArgTyr 298  
 QY 544 GAGCGGTGACGACGACATTTGCTTCGTGGGGCGGCTTCCTTCCTTCCTTCCTTCCTTCCTTC 603  
 Db 299 GluThrValGlnAspCysAsnCysSerIleTyrProGlyHisValSerGlyHisArgMet 318  
 QY 604 GCTTGGATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 627  
 Db 319 AlaTrpAspMetMetMetMetMetMetMetMetMetMetMetMetMetMetMetMetMet 326

RESULT 5  
 GNMVTW

genome polypeptide - hepatitis C virus (strain J)  
 N:Contains: capsid protein C; envelope protein M; major envelope protein E; nonstruct  
 protein NS4A; nonstructural protein NS4B; nonstructural protein NS5  
 C:Species: hepatitis C virus  
 C:Date: 30-Jun-1992 #sequence\_revision 30-Jun-1992 #text\_change 19-Jan-2001  
 C:Accession: A39253; PS0086  
 R:Kato, N.; Hijioka, M.; Ootsuyama, Y.; Nakagawa, M.; Ohkoshi, S.; Sugimura, T.; Shi  
 Proc. Natl. Acad. Sci. U.S.A. 87, 9524-9528, 1990  
 A:Title: Molecular cloning of the human hepatitis C virus genome from Japanese patien  
 A:Reference number: A39253; MUID:91088550; PMID:2175903  
 A:Accession: A39253  
 A:Molecule type: genomic RNA  
 A:Residues: 1-3010 <KAT>  
 A:Cross-references: GB:D90208; NID:9221610; PIDN:BA14233.1; PID:9221611  
 R:Kato, N.; Ohkoshi, S.; Shimotohno, K.  
 Proc. Jpn. Acad. 65B, 219-223, 1989  
 A:Title: Japanese isolates of the non-A, non-B hepatitis viral genome show sequence v  
 A:Reference number: PS0086  
 A:Accession: PS0086  
 A:Molecule type: genomic RNA  
 A:Residues: 2650-2707 <KAT>  
 A:Experimental source: Japanese isolate  
 C:Comment: The cleavage sites of this polypeptide have not been determined.  
 C:Superfamily: hepatitis C virus genome polypeptide  
 C:Keywords: ATP; glycoprotein; hydrolase; nucleotide binding; P-loop; polypeptide; se  
 F:2-115/Product: capsid protein C #status predicted <CPC>  
 F:116-191/Product: envelope protein M #status predicted <EPM>

F:192-389/Product: major envelope protein E #status predicted <ME>  
F:390-729/Product: nonstructural protein NS1 #status predicted <NS1>  
F:730-1006/Product: nonstructural protein NS2 #status predicted <NS2>  
F:1007-1615/Product: hepatitis C virus genome from a single Japanese carrier  
F:1230-1337/Region: nucleotide-binding motif A (P-loop)  
F:1312-1317/Region: nucleotide-binding motif B  
F:1316-1319/Region: DEXH motif  
F:1616-1862/Product: nonstructural protein NS4 #status predicted <NA>  
F:1863-2013/Product: nonstructural protein NS5 #status predicted <NA>  
F:196-209,234,250,305,325,417,423,430,448,532,556,576,623,645,1213,1255,2041,2077,2240,2

## Alignment Scores:

Pred. No.:	1,47e-90	Length:	3010
Score:	1073.00	Matches:	198
Percent Similarity:	98.56%	Conservative:	7
Best Local Similarity:	95.19%	Mismatches:	3
Query Match:	92.90%	Indels:	0
DB:	1	Gaps:	0

US-09-899-303a-7 (1-633) x GMMVCS (1-3010)

QY 4 TTGGGTAAGGTCAATGATACCTTACGTCGCGCTTCGCCGACCTCATGGGGTACATTCCG 63  
DB 119 LeuGlyLysValIleAspThrLeuThrCysGlyPheAlaAspLeuMetGlyTyrIlePro 138  
QY 64 CTCGTCGGCGCCGCCCTAGGGGGCTGCTCCAGAGCCCTGGCGATGGCGGTCTG 123  
DB 139 LeuValIGlyAlaProLeuGlyGlyAlaAlaArgAlaLeuAlaHisGlyValArgValIleu 158  
QY 124 GAAGAGGGGTGAATATGCAACAGGAAATTTGCTGCTTTCCTTATCTTCCTCC 183  
DB 159 GluAspGlyValAsnThrIleThrGlyAsnLeuProGlyCysSerPheSerIlePheLeu 178  
QY 184 TTGGCTTTACTGTCTGTGACCAATTCGACCTTCGCTTATGAGTGCGCAAGCTGCC 243  
DB 179 LeuAlaLeuLeuSerCysLeuThrIleProAlaSerAlaTyrGlyValArgAsnValSer 198  
QY 244 GGGATGACCATGTGTCAGCAAGACAGCTCTCCAACTCAAGATTTGTATGAGCGAGGAC 303  
DB 199 GlyIleTyrHisValThrAsnAspCysSerAsnSerSerIleValIleGluAlaAlaAsp 218  
QY 304 ATGATCATGCACACCCCGGGTGCCTGCGCTTCGGGAGAACAACTCTTCCCGCTGC 363  
DB 219 MetIleMetHisThrProGlyCysValProCysValArgGlnSerAsnPheSerIlePheLeu 238  
QY 364 TTGGAGCGCTCAACCCCAAGCTGCGAGCTAGCAAGCGCAGCTCCCACTAGACAAATA 423  
DB 239 TrpValAlaLeuThrProThrIleuAlaAlaArgAsnSerSerIleProThrThrIle 258  
QY 424 CGAGCCGACGTGATTTGCTCGTTGGGGGCGCTGTTTCTGTTCCGCTATGACGTGGG 483  
DB 259 ArgArgHisValAsnLeuValGlyAlaAlaAlaLeuCysSerAlaMetIleValGly 278  
QY 484 GATCTGAGGAGATCTGCTTCCTGCGTCGCTCCAGCTGATCCATTCGCGCTCCCGGCAAT 543  
DB 279 AspLeuCysGlySerValPheLeuValSerGlnLeuPheThrPheSerProAlaGlyTyr 298  
QY 544 GAGAGCGTCAGAGCTGCAATTCATATCCGCGCACATACACCTTACGCTATG 603  
DB 299 GluThrValGlnAspCysAsnSerIleTyrProGlyHisValSerGlyHisAlaGlyMet 318  
QY 604 GCTTGGGATGATGATGATGATG 627  
DB 319 AlaTrpAspMetMetAsnTrp 326

## RESULT 6

genome polypeptide - hepatitis C virus (strain J1)  
N:Contains: capsid protein C; envelope protein M; hepatitis C virus (EC 3.4.21.98) (nonstructu  
C:Species: hepatitis C virus  
C:Date: 19-May-2000 #sequence\_revision 19-May-2000 #text\_change 19-Jan-2001

C:Accession: A45573  
R:Ranka, T.; Kato, N.; Nakagawa, M.; Ootsuyama, Y.; Cho, M.J.; Nakazawa, T.; Hijikata  
Virus Res. 23, 39-53, 1992  
A:Title: Molecular cloning of hepatitis C virus genome from a single Japanese carrier  
A:Reference number: A45573; MUID:92295714; PMID:1318627  
A:Accession: A45573  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-3010 <TRAN>  
A:Cross-References: GB:DI1168; GB:DO1171; NID:g221612; PIDN:BA01943.1; PID:g221613  
A:Experimental source: HCV-JT  
A:Note: sequence extracted from NCBI backbone (NCBI:106206, NCBI:106207)  
C:Superfamily: hepatitis C virus genome polypeptide  
C:Keywords: ATP; glycoprotein; hydrolase; nucleotide binding; P-loop; polypeptide; se  
F:2-115/Product: capsid protein C #status predicted <CPC>  
F:116-191/Product: envelope protein M #status predicted <EPM>  
F:192-389/Product: major envelope protein E #status predicted <ME>  
F:390-729/Product: nonstructural protein NS1 #status predicted <NS1>  
F:730-1006/Product: nonstructural protein NS2 #status predicted <NS2>  
F:1007-1615/Product: hepatitis C virus genome polypeptide  
F:1230-1337/Region: nucleotide-binding motif A (P-loop)  
F:1312-1317/Region: nucleotide-binding motif B  
F:1316-1319/Region: DEXH motif  
F:1616-1862/Product: nonstructural protein NS4 #status predicted <NA>  
F:1863-2013/Product: nonstructural protein NS5 #status predicted <NA>  
F:2014-3010/Product: nonstructural protein NS5 #status predicted <NS5>  
Alignment Scores:  
Pred. No.: 1,47e-90 Length: 3010  
Score: 1073.00 Matches: 197  
Percent Similarity: 98.08% Conservative: 7  
Best Local Similarity: 94.71% Mismatches: 4  
Query Match: 92.90% Indels: 0  
DB: 1 Gaps: 0  
US-09-899-303a-7 (1-633) x A45573 (1-3010)  
QY 4 TTGGGTAAGGTCAATGATACCTTACGTCGCGCTTCGCCGACCTCATGGGGTACATTCCG 63  
DB 119 LeuGlyLysValIleAspThrLeuThrCysGlyPheAlaAspLeuMetGlyTyrIlePro 138  
QY 64 CTCGTCGGCGCCGCCCTAGGGGGCTGCTCCAGAGCCCTGGCGATGGCGGTCTG 123  
DB 139 LeuValIGlyAlaProLeuGlyGlyAlaAlaArgAlaLeuAlaHisGlyValArgValIleu 158  
QY 124 GAAGAGGGGTGAATATGCAACAGGAAATTTGCTGCTTTCCTTATCTTCCTCC 183  
DB 159 GluAspGlyValAsnThrIleThrGlyAsnLeuProGlyCysSerPheSerIlePheLeu 178  
QY 184 TTGGCTTTACTGTCTGTGACCAATTCGACCTTCGCTTATGAGTGCGCAAGCTGCC 243  
DB 179 LeuAlaLeuLeuSerCysLeuThrIleProAlaSerAlaTyrGlyValArgAsnValSer 198  
QY 244 GGGATGACCATGTGTCAGCAAGACAGCTCTCCAACTCAAGATTTGTATGAGCGAGGAC 303  
DB 199 GlyIleTyrHisValThrAsnAspCysSerAsnSerSerIleValIleGluAlaAlaAsp 218  
QY 304 ATGATCATGCACACCCCGGGTGCCTGCGCTTCGGGAGAACAACTCTTCCCGCTGC 363  
DB 219 MetIleMetHisThrProGlyCysValProCysValArgGlnSerAsnPheSerIlePheLeu 238  
QY 239 TrpValAlaLeuThrProThrIleuAlaAlaArgAsnSerSerIleProThrThrIle 258  
QY 424 CGAGCCGACGTGATTTGCTCGTTGGGGGCGCTGTTTCTGTTCCGCTATGACGTGGG 483  
DB 259 ArgArgHisValAsnLeuValGlyAlaAlaAlaPheCysSerAlaMetIleValGly 278  
QY 484 GATCTGAGGAGATCTGCTTCCTGCGTCGCTCCAGCTGATCCATTCGCGCTCCCGGCAAT 543  
DB 279 AspLeuCysGlySerValPheLeuValSerGlnLeuPheThrPheSerProAlaGlyTyr 298

OY 544 GAGAGGTCGACGACTGCAATTCGTCATCTCTCCCGGACATATACAGATACCGGTATG 603  
 |||||||  
 Db 299 GTGTTTATGTAAGACGTCATGTCATGTCATGTCATGTCATGTCATGTCATGTCATG 318  
 |||||||  
 OY 604 GCTTGAGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 627  
 |||||||  
 Db 319 AATATPAP 326  
 |||||||  
 RESULT 7  
 S18030  
 genome polypeptide - hepatitis C virus (Isolate JKI)  
 N:Contains: capsid protein C; envelope protein M; hepatitis C virus (EC 3.4.21.98) (nonstruct  
 protein NS4a; nonstructural protein NS4b; nonstructural protein NS5  
 C;Species: hepatitis C virus  
 A:Variety: Isolate JKI  
 C:Accession: S18030  
 C:Date: 19-May-2000  
 C:Revision: 19-May-2000  
 R:Honda, M.; Kaneo, S.; Masashi, U.; Kobayashi, K.; Murakami, S.  
 submitted to the EMBL Data Library, September 1991  
 A:Description: A whole genome of hepatitis C virus cDNA was isolated from a single patie  
 A:Reference number: S18028  
 A:Accession: S18030  
 A:Molecule type: genomic RNA  
 A:Residues: 1-3010 <HOW>  
 A:CROSS-references: EMBL:X61596; NID:959478; PIDN:CAA43793.1; PID:959479  
 A:Experimental source: Isolate JKI from an individual  
 R:Honda, M.; Kaneo, S.; Uncura, M.; Kobayashi, K.; Murakami, S.  
 Arch. Virol. 128, 163-169, 1993  
 A:Title: Sequence analysis of putative structural regions of hepatitis C virus isolated  
 A:Reference number: A48332; MUID:93119270; PMID:8380322  
 A:Accession: S33570  
 A:Molecule type: genomic RNA  
 A:Residues: 1-547, 'T', 549-621, 'V', 623-624, 'S', 626-652, 'DL', 655-761, 'T', 763-782 <HOW>  
 A:CROSS-references: EMBL:X61591  
 A:Note: this sequence is inconsistent with the nucleotide translation  
 as Trp, and Trp for residue 771 as Ser  
 A:Note: sequence extracted from NCBI backbone (NCBI:121747, NCBI:121748)  
 C:Superfamily: hepatitis C virus genome polypeptide  
 C:Keywords: ATP; glycoprotein; hydrolyase; nucleotide binding; P-loop; polypeptide; serine  
 F:2-115/Product: capsid protein C #status predicted <CPC>  
 F:116-101/Product: envelope protein M #status predicted <EPM>  
 F:192-389/Product: major envelope protein E #status predicted <MEP>  
 F:390-729/Product: nonstructural protein NS1 #status predicted <NS1>  
 F:730-1006/Product: nonstructural protein NS2 #status predicted <NS2>  
 F:1007-1615/Product: hepatitis C virus NS3 #status predicted <NS3>  
 F:1230-1237/Product: nucleotide-binding motif A (P-loop)  
 F:1312-1317/Product: nucleotide-binding motif B  
 F:1316-1319/Product: DEXH motif  
 F:1616-1862/Product: nonstructural protein NS4a #status predicted <NS4a>  
 F:1863-2013/Product: nonstructural protein NS4b #status predicted <NS4b>  
 F:2014-3010/Product: nonstructural protein NS5 #status predicted <NS5>  
 F:196,209,224,250,305,417,423,448,532,540,556,576,623,645/Binding site: carbohydrate (A

```

Db      159 GUAAGGlyValAsnTyrAlaThrGlYasnLeuProGlyCysSerPheSerIlePheHeu 178
|||||
Db      184 TTGGCTTAACGTCTCGTCCTGCACATTCACAGTTCCGGTTATGAGGTGGCAAGTGCC 243
|||||
Db      179 LeuAlaLeuLeuSerCysLeuThrValProValSerTrpTyrGluValAlaArgAsnValSer 198
|||||
Qy      244 GGAGATGACCATGTCACAGAACGACTGTCCAATCCAACTATTGTATGAGCAGCGAC 303
|||||
Db      199 GlyValTyrHisValThrAsnApsCysSerAsnSerIleValTyrGluAlaAlaAsp 218
|||||
Qy      304 ATGATCATGCACACACCCCCGGGTGGCTGCCCTGCGTTCCGGAGAACACTCTTCCGCTGC 363
|||||
Db      219 MetIleMetHisThrProGlyCysValProCysValArgGluGlyAsnSerSerArgCys 238
|||||
Qy      364 TGGGTAAGGCGTCACCCCACAGCTGCACAGTGAAGAACGACAGGCGCCCACTACGACATA 423
|||||
Db      239 TrpValAlaLeuThrProThrLeuAlaAlaArgAsnSerSerIleProThrThrIle 258
|||||
Qy      424 CGACGSCACGTCGATTTGCTGCTGGGGCGGCGCTTTCTGTCCGCTATGATAGCTGGGG 483
|||||
Db      259 ArgArgHisValAspLeuLeuValGlyAlaAlaAlaLeuCysSerAlaMetTyrValGly 278
|||||
Qy      484 GATCTCTCGGATCTGTCTTCTGCTGCTCCACAGCTGTTCACACATCGCCCTCCGGCAT 543
|||||
Db      279 AspLeuCysGlySerValPheLeuValSerGlnLeuPheThrPheSerProArgArgTyr 298
|||||
Qy      544 GAGACGGTGCAGGAGTGGAAATTGCTCATTCATCCGGGCGCACATACAGAGTACCAGTATG 603
|||||
Db      299 GluThrValGlnAspCysAsnCysSerLeuTyrProGlyHisValSerGlyHisArgMet 318
|||||
Qy      604 GCTTGGAATGATGATGATGACATGG 627
|||||
Db      319 AlaTrpAspMetMetMetLsnTrp 326
|||||

RESULT 8
SI18031
genome polypotein - hepatitis C virus (isolate JK2) (fragment)
N.Contains: core protein; envelope protein 1; nonstructural protein 2; NSL/E2 protein
C.Species: hepatitis C virus
A.Variety: isolate JK2
C.Date: 30-Jun-1992 #sequence_revision 30-Jun-1992 #text_change 17-Nov-2000
C.Accession: SI18031
R.Honda, M.; Kaneo, S.; Masashi, U.; Kobayashi, K.; Murakami, S.
submitted to the EMBL Data Library, September 1991
A.Description: Sequence analysis of putative structural regions of Hepatitis C Virus
A.Reference number: SI18029
A.Accession: SI18031
A.Molecule type: genomic RNA
A.Residues: 1-782 <HON>
A.Cross-references: EMBL:X61593
C.Experimental source: isolate JK2
C.Superfamily: hepatitis C virus genome polypotein
C.Keywords: capsid protein; core protein; envelope protein; glycoprotein; nonstructur
F.1-191/Product: core protein #status predicted <MAT1>
F.192-383/Product: envelope protein 1 #status predicted <MAT2>
F.384-733/Product: NSL/E2 protein #status predicted <MAT3>
F.734-782/Product: nonstructural protein 2 (fragment) #status predicted <MAT4>

Alignment Scores:
Pred. No.:          7,7e+90          Length:          782
Score:              1065.00         Matches:           196
Percent Similarity: 98.08%         Conservative:       8
Best Local Similarity: 94.23%        Mismatches:        4
Query Match:        92.21%          Indels:            0
DB:                  2               Gaps:              0

US-09-839-303A-7 (1-633) x SI18031 (1-782)
Qy      4 TTGGTAGGATGATGATGATACCTTACGCGGGCTTCGCGGACATCATGGGTATCTCCG 63
|||||
Db      119 LeugLyVsvalIlleasPrhleuThrcysgslYpealaaSprleumetGlyTyrIlepro 138
|||||

```

```

QY 64 CTGTCGGCGCCCGCCCTAGGGGGTGCCTCCAGAGCCCTGGCGCATGGCGTCCGGTCTG 123
    |||
Db 139 LeuValGlyAlaProLeuGlyGlyAlaAlaArgAlaLeuAlaHisGlyValArgValLeu 158
QY 124 GAAGACGGCGCTGAACATGCAACAGGGAATTTCCTGGTTCCTCTTCTATCTTCCTC 183
    |||
Db 159 GluAspGlyValAsnThrAlaThrGlyAsnLeuProGlySerPheSerIlePheLeu 178
QY 184 TTGGCTTTACTGCTGCTGTGACCAATTCAGCTTCGCTTATGAGTGCGCAACGTGCC 243
    |||
Db 179 LeuAlaLeuLeuSerCysLeuThrIleProAlaSerAlaTyrGluValArgAsnValSer 198
QY 244 GGGATGACCATGTCAGAGACGACGCTGCACACTCAAGCATTTGTATGAGCAGCGGAC 303
    |||
Db 199 GlyLeuThrHisValThrAsnAspCysSerAsnSerSerIleValTyrGluAlaAlaAsp 218
QY 304 ATGATCATGCACACCCCGGGTGCCTGCGCTTCGCGGAGAACACTTCCTCCGCTGC 363
    |||
Db 219 LeuIleMetHisThrProGlyCysValProCysValLeuGluLysAsnSerSerArgCys 238
QY 364 TTGGTAGCGCTCAACCCCGGCTGCGCTGCGCTTCGCGGAGAACACTTCCTCCGCTGC 423
    |||
Db 239 TrpValAlaLeuThrProThrLeuAlaAlaArgAsnSerSerIleProThrThrIle 258
QY 424 CGAGCGCAGCTGCATTTGCTCGCTGGGGGGCGCTTTCGTTCCGCTATGACGTGGG 483
    |||
Db 259 ArgArgHisValAspLeuLeuValGlyAlaAlaAlaPheCysSerAlaMetTyrValGly 278
QY 484 GATCTGCGGATGCTGCTCTCTGCTCCAGCTGTCACATTCGCGCTTCGCGGAT 543
    |||
Db 279 AspLeuCysGlySerValPheLeuValSerGlnLeuPheThrPheSerProAlaArgTyr 298
QY 544 GAGACGGTCGAGAGTCGCAATTGCTCAATCTATCCGGCCATCAACAGGTCAACCGTATG 603
    |||
Db 299 GluThrValGlnAspCysAsnCysSerLeuTyrSerGlyHisValSerGlyHisArgMet 318
QY 604 GCTTGGATATGATGATGAACCTGG 627
    |||
Db 319 AlaTrpAspMetMetMetAsnTrp 326

```

## RESULT 9

S19876  
 genome polyprotein - hepatitis C virus (isolate JK5) (fragment)  
 N:contains: core protein; envelope protein 1; nonstructural protein 2; NS1/E2 protein  
 C:Species: hepatitis C virus  
 A:Variety: isolate JK5  
 C:Date: 30-Jun-1992 #sequence\_revision 30-Jun-1992 #text\_change 17-Nov-2000  
 C:Accession: S19876  
 R:Honda, M.; Kaneko, S.; Masashi, U.; Kobayashi, K.; Murakami, S.  
 submitted to the EMBL Data Library, September 1991  
 A:Description: Sequence analysis of putative structural regions of Hepatitis C Virus iso  
 A:Reference number: S18029  
 A:Accession: S19876  
 A:Molecule type: genomic RNA  
 A:Residues: 1-782 <HON>  
 A:Cross-references: EMBL:X61595; NID:g59486; PIDN:CAA43792.1; PID:g59487  
 A:Experimental source: isolate JK5  
 C:Superfamily: hepatitis C virus genome polyprotein  
 C:Keywords: capsid protein; core protein; envelope protein; glycoprotein; nonstructural  
 F:1-191/Product: core protein #status predicted <MAT1>  
 F:192-383/Product: envelope protein 1 #status predicted <MAT2>  
 F:384-733/Product: NS1/E2 protein #status predicted <MAT3>  
 F:734-782/Product: nonstructural protein 2 (fragment) #status predicted <MAT4>

## Alignment Scores:

Pred. No.:	1,18e-89	Length:	782
Score:	1063.00	Matches:	196
Percent Similarity:	97.60%	Conservative:	7
Best Local Similarity:	94.23%	Mismatches:	5
Query Match:	92.03%	Indels:	0
DB:	2	Gaps:	0

US-09-899-303A-7 (1-633) x S19876 (1-782)

```

QY 4 TTGGGTAAGGTCAATGATACCTTACGTGGCGCTTCGCCAGCCTCATGGGTACATTCCG 63
    |||
Db 119 LeuGlyValIleAspThrLeuThrCysGlyPheAlaAspProMetGlyTyrIlePro 138
QY 64 CTGTCGGCGCCCGCCCTAGGGGGTGCCTCCAGAGCCCTGGCGCATGGCGTCCGGTCTG 123
    |||
Db 139 LeuValGlyAlaProLeuGlyGlyAlaAlaArgAlaLeuAlaHisGlyValArgValLeu 158
QY 124 GAAGACGGCGCTGAACATGCAACAGGGAATTTCCTGGTTCCTCTTCTATCTTCCTC 183
    |||
Db 159 GluAspGlyValAsnThrAlaThrGlyAsnLeuProGlySerPheSerIlePheLeu 178
QY 184 TTGGCTTTACTGCTGCTGTGACCAATTCAGCTTCGCTTATGAGTGCGCAACGTGCC 243
    |||
Db 179 LeuAlaLeuLeuSerCysLeuThrIleProAlaSerAlaTyrGluValArgAsnValSer 198
QY 244 GGGATGACCATGTCAGAGACGACGCTGCACACTCAAGCATTTGTATGAGCAGCGGAC 303
    |||
Db 199 GlyLeuThrHisValThrAsnAspCysSerAsnSerSerIleValTyrGluAlaAlaAsp 218
QY 304 ATGATCATGCACACCCCGGGTGCCTGCGCTTCGCGGAGAACACTTCCTCCGCTGC 363
    |||
Db 219 LeuIleMetHisThrProGlyCysValProCysValLeuGluLysAsnSerSerArgCys 238
QY 364 TTGGTAGCGCTCAACCCCGGCTGCGCTGCGCTTCGCGGAGAACACTTCCTCCGCTGC 423
    |||
Db 239 TrpValAlaLeuThrProThrLeuAlaAlaArgAsnSerSerIleProThrThrIle 258
QY 424 CGAGCGCAGCTGCATTTGCTCGCTGGGGGGCGCTTTCGTTCCGCTATGACGTGGG 483
    |||
Db 259 ArgArgHisValAspLeuLeuValGlyAlaAlaAlaPheCysSerAlaMetTyrValGly 278
QY 484 GATCTGCGGATGCTGCTCTCTGCTCCAGCTGTCACATTCGCGCTTCGCGGAT 543
    |||
Db 279 AspLeuCysGlySerValPheLeuValSerGlnLeuPheThrPheSerProAlaArgTyr 298
QY 544 GAGACGGTCGAGAGTCGCAATTGCTCAATCTATCCGGCCATCAACAGGTCAACCGTATG 603
    |||
Db 299 GluThrValGlnAspCysAsnCysSerLeuTyrSerGlyHisValSerGlyHisArgMet 318
QY 604 GCTTGGATATGATGATGAACCTGG 627
    |||
Db 319 AlaTrpAspMetMetMetAsnTrp 326

```

## RESULT 10

S18032  
 genome polyprotein - hepatitis C virus (isolate JK4) (fragment)  
 N:contains: core protein; envelope protein 1; nonstructural protein 2; NS1/E2 protein  
 C:Species: hepatitis C virus  
 A:Variety: isolate JK4  
 C:Date: 30-Jun-1992 #sequence\_revision 30-Jun-1992 #text\_change 17-Nov-2000  
 C:Accession: S18032  
 R:Honda, M.; Kaneko, S.; Masashi, U.; Kobayashi, K.; Murakami, S.  
 submitted to the EMBL Data Library, September 1991  
 A:Description: Sequence analysis of putative structural regions of Hepatitis C Virus  
 A:Reference number: S18029  
 A:Accession: S18032  
 A:Molecule type: genomic RNA  
 A:Residues: 1-782 <HON>  
 A:Cross-references: EMBL:X61594  
 A:Experimental source: isolate JK4  
 C:Superfamily: hepatitis C virus genome polyprotein  
 C:Keywords: capsid protein; core protein; envelope protein; glycoprotein; nonstructur  
 F:1-191/Product: core protein #status predicted <MAT1>  
 F:192-383/Product: envelope protein 1 #status predicted <MAT2>  
 F:384-733/Product: NS1/E2 protein #status predicted <MAT3>  
 F:734-782/Product: nonstructural protein 2 (fragment) #status predicted <MAT4>

## Alignment Scores:

Pred. No.:	1,18e-89	Length:	782
Score:	1063.00	Matches:	194
Percent Similarity:	97.12%	Conservative:	8

Best Local Similarity: 93.27%  
Query Match: 92.03%  
DB: 2

Mismatches: 6  
Indels: 0  
Gaps: 0

US-09-899-303a-7 (1-633) x S18032 (1-782)

QY 4 TTGGGTAAAGTCATGATACCCCTTACGTGGGGCTTCCGCCATCATGGGTACATTCCG 63  
DB 119 LeuGLySVaLIleAPRrLLeuThrCysGLyPheAlaSPLeuMetGLyTrILePro 138  
QY 64 CTCGTCCGGCCCCCCTAGGGGGTGTCCAGACCCCTGGGCGCATGGCGTCCGGTCTG 123  
DB 139 LeuValGLyAlaProLeuGLyGLyValAlaAlaArgAlaLeuAlaHisGLyValAlaLeu 158  
QY 124 GAAGACGGCGTGAACATAAGCAAGGAAATTTGCCGTGTCTCTCTATCTTCTC 183  
DB 159 GluASPGLyAlaSNrYrAlaThrGLyASNLeuProGLySSerPheSerILePheLeu 178  
QY 184 TTGGCTTACTGTCTCTGTGACCATTCAGCTTCCGCTTATGAGTCCGACAGTGTCC 243  
DB 179 LeuAlaLeuLeuSerCysLeuThrValProValSerThrTyrgluValAlaArgASNValSer 198  
QY 244 GGGATGACCATGTCACAGAACGACTGCTCCAACTCAAGCATTTGTATGAGGCGAGCGGAC 303  
DB 199 GlyValIyrHisValIThrASNAsPCysSerASNserSerILeValTyrgluAlaAlaASP 218  
QY 304 ATGATCATGACACACCCCGGGTGGCTGCTCCGCTTCCGGAGAACACTCTTCCCGTGC 363  
DB 219 MetIleMetHisThrProGLyCysValProCysValAlaArgGLuASNserSerArgCys 238  
QY 364 TGGGTACCGCTCACCCCGGCTGCGAGCTAGAACCCGACCGTCCCACTACGACATA 423  
DB 239 TrpValAlaLeuThrProThrLeuAlaAlaArgASNserSerILeProThrThrIle 258  
QY 424 CGAGCCACGTCGATTTGCTGCTGGGGGGGCTTCTTCTTCCGATGATGAGTGGGG 483  
DB 259 ArgArgHisValAspLeuLeuValGLyAlaAlaAlaLeuCysSerAlaMetTyValGLy 278  
QY 484 GATCTCTGCGGANTGTCTCTCTCTCTCCAGCTGTTCACCATCTGCGCTCCGCGGAT 543  
DB 279 AspLeuCysGLySerValIlePheLeuValSerGlnLeuPheThrPheSerProArgArgTy 298  
QY 544 GAGACGGTGCAGACTGCATATGCTCAATCTATCCCGGCACATACAGGTACACCGTATG 603  
DB 299 GluThrValGlnAspCysAsnCysSerLeuTyProGlnHisValSerGlnHisArgMet 318  
QY 604 GCTTGGCATATGATGATGAAGTGG 627  
DB 319 AlaTrpAspMetMetMetAsnTrp 326

RESULT 11

S21471  
genome polyprotein - hepatitis C virus (fragment)  
N:Contains: capsid protein; envelope protein  
C:Species: hepatitis C virus  
C>Date: 20-Feb-1995 #sequence\_revision 19-Apr-1996 #text\_change 17-Nov-2000  
C:Accession: S21471  
R:Mogam, W.K.  
submitted to the EMBL Data Library, April 1992  
A:Reference number: S21471  
A:Accession: S21471  
A:Molecule type: genomic RNA  
A:Residues: 1-369 <NOG>  
A:Cross-references: EMBL:X65924; NID:959466; PIDN:CAA6717.1; PID:959467  
C:Superfamily: hepatitis C virus genome polyprotein  
C:Keywords: capsid protein; envelope protein; glycoprotein; polyprotein

Alignment Scores:  
Pred. No.: 2.17e-89  
Score: 1060.00  
Percent Similarity: 97.60%  
Best Local Similarity: 94.71%  
Query Match: 91.77%

DB: 2 Gaps: 0

US-09-899-303a-7 (1-633) x S21471 (1-369)

QY 4 TTGGGTAAAGTCATGATACCCCTTACGTGGGGCTTCCGCCATCATGGGTACATTCCG 63  
DB 119 LeuGLySVaLIleAPRrLLeuThrCysGLyPheAlaSPLeuMetGLyTrILePro 138  
QY 64 CTCGTCCGGCCCCCCTAGGGGGTGTCCAGACCCCTGGGCGCATGGCGTCCGGTCTG 123  
DB 139 LeuValGLyAlaProLeuGLyGLyValAlaAlaArgAlaLeuAlaHisGLyValAlaLeu 158  
QY 124 GAAGACGGCGTGAACATAAGCAAGGAAATTTGCCGTGTCTCTCTATCTTCTC 183  
DB 159 GluASPGLyAlaSNrYrAlaThrGLyASNLeuProGLySSerPheSerILePheLeu 178  
QY 184 TTGGCTTACTGTCTCTGTGACCATTCAGCTTCCGCTTATGAGTCCGACAGTGTCC 243  
DB 179 LeuAlaLeuLeuSerCysLeuThrValProValSerThrTyrgluValAlaArgASNValSer 198  
QY 244 GGGATGACCATGTCACAGAACGACTGCTCCAACTCAAGCATTTGTATGAGGCGAGCGGAC 303  
DB 199 GlyValIyrHisValIThrASNAsPCysSerASNserSerILeValTyrgluAlaAlaASP 218  
QY 304 ATGATCATGACACACCCCGGGTGGCTGCTCCGCTTCCGGAGAACACTCTTCCCGTGC 363  
DB 219 ThrIleMetHisThrProGLyCysValProCysValAlaArgGLuASNserSerArgCys 238  
QY 364 TGGGTACCGCTCACCCCGGCTGCGAGCTAGAACCCGACCGTCCCACTACGACATA 423  
DB 239 TrpValAlaLeuThrProThrLeuAlaAlaArgASNserSerILeProThrThrIle 258  
QY 424 CGAGCCACGTCGATTTGCTGCTGGGGGGGCTTCTTCTTCCGATGATGAGTGGGG 483  
DB 259 ArgArgHisValAspLeuLeuValGLyAlaAlaAlaLeuCysSerAlaMetTyValGLy 278  
QY 484 GATCTCTGCGGANTGTCTCTCTCTCTCCAGCTGTTCACCATCTGCGCTCCGCGGAT 543  
DB 279 AspLeuCysGLySerValIlePheLeuValSerGlnLeuPheThrPheSerProArgArgTy 298  
QY 544 GAGACGGTGCAGACTGCATATGCTCAATCTATCCCGGCACATACAGGTACACCGTATG 603  
DB 299 GluThrValGlnAspCysAsnCysSerLeuTyProGlnHisValSerGlnHisArgMet 318  
QY 604 GCTTGGCATATGATGATGAAGTGG 627  
DB 319 AlaTrpAspMetMetMetAsnTrp 326

RESULT 12

GNMVTTC  
genome polyprotein - hepatitis C virus  
N:Contains: capsid protein C; envelope protein M; hepacivirin (PC 3.4.21.98) (nonstru  
protein NS4a; nonstructural protein NS4b; nonstructural protein NS5  
C:Species: hepatitis C virus  
C>Date: 31-Mar-1992 #sequence\_revision 31-Mar-1992 #text\_change 19-Jan-2001  
C:Accession: A38465  
R:Takamizawa, A.; Mori, C.; Fuke, I.; Manabe, S.; Murakami, S.; Fujita, J.; Onishi, E  
J. Virol. 65, 1105-1113, 1991  
A:Title: Structure and organization of the hepatitis C virus genome isolated from hum  
A:Reference number: A38465; M01D:91140696; PMID:18477440  
A:Accession: A38465  
A:Molecule type: genomic RNA  
A:Residues: 1-3010 <TKA>  
A:Cross-references: EMBL:M58335; NID:9329770; PIDN:AA72945.1; PID:9329771  
C:Superfamily: hepatitis C virus genome polyprotein  
C:Keywords: ATP; capsid protein; envelope protein; glycoprotein; hydrolase; nonstruct  
F:2-115/Product: capsid protein C #status predicted <CP>  
F:116-191/Product: envelope protein M #status predicted <EPK>  
F:192-389/Product: major envelope protein E #status predicted <NEE>  
F:390-729/Product: nonstructural protein NS1 #status predicted <NS1>  
F:730-1006/Product: nonstructural protein NS2 #status predicted <NS2>  
F:1007-1615/Product: hepacivirin #status predicted <NS3>  
F:1230-1237/Region: nucleotide-binding motif A (P-loop)

F:1312-1317/Region: nucleotide-binding motif B  
F:1316-1319/Region: DEXH motif  
F:1616-1862/Product: nonstructural protein NS4a #status predicted <N4>  
F:1863-2013/Product: nonstructural protein NS4b #status predicted <N4>  
F:2014-3010/Product: nonstructural protein NS5 #status predicted <NS5>  
F:196,209,234,250,305,325,417,423,430,448,532,540,556,576,623,645,1213,1255,2041,2077,22

## Alignment Scores:

Pred. No.:	2,31e-89	Length:	3010
Score:	1060.00	Matches:	194
Percent Similarity:	97.12%	Conservative:	8
Best Local Similarity:	93.27%	Mismatches:	6
Query Match:	91.77%	Indels:	0
DB:	1	Gaps:	0

US-09-899-303a-7 (1-633) x GMYTC (1-3010)

```
QY 4 TTGGGTAAAGGTCAATGATACCTTACGTCGGGCTTCGCGACCTCAATGGGGTACATTCCG 63
    |||||||
Db 119 LeuGlyValValIleAspThrLeuThrCysGlyPheAlaAspLeuMetGlyTyrIlePro 138
QY 64 CTGCTGGGCGCCCCCTAGGGGGTGTGCTCCAGAGCCCTGGCGATGGCGGTCTCTG 123
    |||||||
Db 139 LeuValGlyAlaProLeuGlyGlyAlaAlaArgAlaLeuAlaHisGlyValArgValLeu 158
QY 124 GAAGAGGCGCTGAACATGCAACAGGGAATTGGCTGGTCTCTCTTCTTATCTTCTC 183
    |||||||
Db 159 GluAspGlyAlaAsnTyrAlaThrGlyAsnLeuProGlyCysSerPheSerIlePheLeu 178
QY 184 TTGGCTTAACTGTCGTCTGACCAATTCGACCTTCCGCTTATAGAGGCGCAACGTTC 243
    |||||||
Db 179 LeuAlaLeuLeuSerCysLeuThrThrProAlaSerAlaTyrGluValHisAsnValSer 198
QY 244 GGGATGTACCATGTCAAGAGCACTGCTCCAACTCAAGCATTTGTATGAGGACGGAC 303
    |||||||
Db 199 GlyIleTyrHisValIleThrAsnAspCysSerAsnAlaSerIleValTyrGluAlaAlaAsp 218
QY 304 ATGATATGACACACCCCGGGGTGCGTCCCTGGCTTCGGAGAAACAATCTTCCCGTGC 363
    |||||||
Db 219 LeuIleMetHisThrProGlyCysValProCysValArgIleGluIleAsnSerSerArgCys 238
QY 364 TGGGTAGCGCTCAACCCCAAGCTCGAGCTGAGAGCAAGCCAGCCGCTCCCACTAGACAATA 423
    |||||||
Db 239 TrpValAlaLeuThrProThrLeuAlaAlaArgAsnValThrIleProThrThrIle 258
QY 424 CGAGCGCAGCTGATTTGCTGTTGGGGCGGCTTCTTCTGTTCCGCTATGTACGTGGG 483
    |||||||
Db 259 ArgArgHisValAspLeuLeuValGlyAlaAlaAlaPheCysSerAlaMetTyrValGly 278
QY 484 GATCTGCGGAGATGCTGTCCTCGTCCCTCCACCTGTTCACCAATCTCCGCTCCCGGAT 543
    |||||||
Db 279 AspLeuCysGlySerValPheLeuValSerGlnLeuPheThrPheSerProArgArgHis 298
QY 544 GAGACGGTGCAGAGTGAATTTGCTCAATCTATCCGCGACATTAACAGGTCAACCGTATG 603
    |||||||
Db 299 ValThrLeuGlnAspCysAsnCySerIleTyrProGlyHisValSerGlyHisArgMet 318
QY 604 GCTTGGGATATGATGATGAACCTGG 627
    |||||||
Db 319 AlaTrpAspMetMetMetAsnTrp 326
```

## RESULT 13

JQ1925  
polyprotein - hepatitis C virus (isolate HCV-KE)  
N:Contains: C protein; E1 protein; E2/NS1 protein  
C:Species: hepatitis C virus  
C:Date: 03-May-1994 #sequence\_revision 03-May-1994 #text\_change 17-Nov-2000  
C:Accession: JQ1925  
R:Abde, K.; Inchauspe, G.; Fujisawa, K.  
J. Gen. Virol. 73, 2725-2729, 1992  
A:Title: Genomic characterization and mutation rate of hepatitis C virus isolated from a  
A:Reference number: JQ1925; MUID:93019030; PMID:1383400  
A:Accession: JQ1925

A:Molecule type: mRNA  
A:Residues: 1-520 <ABE>  
A:Cross-references: DDBJ:D10687; NID:9221544; PDB:BA01529.1; PID:9221545  
C:Superfamily: hepatitis C virus genome polyprotein  
C:Keywords: polyprotein; transmembrane protein  
F:191-/Product: C protein #status predicted <CPR>  
F:192-383/Product: E1 protein #status predicted <E1R>  
F:384-520/Product: E2/NS1 protein #status predicted <E2P>

## Alignment Scores:

Pred. No.:	4,15e-89	Length:	520
Score:	1057.00	Matches:	195
Percent Similarity:	96.63%	Conservative:	6
Best Local Similarity:	93.75%	Mismatches:	7
Query Match:	91.52%	Indels:	0
DB:	2	Gaps:	0

US-09-899-303a-7 (1-633) x JQ1925 (1-520)

```
QY 4 TTGGGTAAAGGTCAATGATACCTTACGTCGGGCTTCGCGACCTCAATGGGGTACATTCCG 63
    |||||||
Db 119 LeuGlyValValIleAspThrLeuThrCysGlyPheAlaAspLeuMetGlyTyrIlePro 138
QY 64 CTGCTGGGCGCCCCCTAGGGGGTGTGCTCCAGAGCCCTGGCGATGGCGGTCTCTG 123
    |||||||
Db 139 LeuValGlyAlaProLeuGlyGlyAlaAlaArgAlaLeuAlaHisGlyValArgValLeu 158
QY 124 GAAGAGGCGCTGAACATGCAACAGGGAATTGGCTGGTCTCTCTTCTTATCTTCTC 183
    |||||||
Db 159 GluAspGlyAlaAsnTyrAlaThrGlyAsnLeuProGlyCysSerPheSerIlePheLeu 178
QY 184 TTGGCTTAACTGTCGTCTGACCAATTCGACCTTCCGCTTATAGAGGCGCAACGTTC 243
    |||||||
Db 179 SerAlaLeuMetSerCysLeuThrThrProAlaSerAlaTyrGluValArgAsnValSer 198
QY 244 GGGATGTACCATGTCAAGAGCACTGCTCCAACTCAAGCATTTGTATGAGGACGGAC 303
    |||||||
Db 199 GlyIleTyrHisValIleThrAsnAspCysSerAsnSerIleAlaTyrGluAlaAlaGly 218
QY 304 ATGATATGACACACCCCGGGGTGCGTCCCTGGCTTCGGAGAAACAATCTTCCCGTGC 363
    |||||||
Db 219 MetIleMetHisThrProGlyCysValProCysValArgIleGluAsnAsnSerSerArgCys 238
QY 364 TGGGTAGCGCTCAACCCCAAGCTCGAGCTGAGAGCAAGCCAGCCGCTCCCACTAGACAATA 423
    |||||||
Db 239 TrpValAlaLeuThrProThrLeuAlaAlaArgAsnAlaSerValProThrThrIle 258
QY 424 CGAGCGCAGCTGATTTGCTGTTGGGGCGGCTTCTTCTGTTCCGCTATGTACGTGGG 483
    |||||||
Db 259 ArgArgHisValAspLeuLeuValGlyAlaAlaAlaThrLeuCysSerAlaMetTyrValGly 278
QY 484 GATCTGCGGAGATGCTGTCCTCGTCCCTCCACCTGTTCACCAATCTCCGCTCCCGGAT 543
    |||||||
Db 279 AspLeuCysGlySerValPheLeuValSerGlnLeuPheThrPheSerProArgArgHis 298
QY 544 GAGACGGTGCAGAGTGAATTTGCTCAATCTATCCGCGACATTAACAGGTCAACCGTATG 603
    |||||||
Db 299 GluThrValGlnAspCysAsnCySerIleTyrProGlyHisValSerGlyHisArgMet 318
QY 604 GCTTGGGATATGATGATGAACCTGG 627
    |||||||
Db 319 AlaTrpAspMetMetMetAsnTrp 326
```

## RESULT 14

JH0711  
genome polyprotein - hepatitis C virus (strain PKC1) (fragments)  
N:Contains: envelope protein E1; envelope protein E2; nonstructural protein NS1; nons  
C:Species: hepatitis C virus  
C:Date: 17-Apr-1993 #sequence\_revision 17-Apr-1993 #text\_change 17-Nov-2000  
C:Accession: JH0711  
R:Lin, K.; Hu, Z.; Li, H.; Prince, A.M.; Inchauspe, G.  
Gene 114, 245-250, 1992  
A:Title: Genomic typing of hepatitis C viruses present in China.

A:Reference number: JH0711; MUID:92290283; PMID:1318245  
A:Accession: JH0711  
A:Molecule type: genomic RNA  
A:Residues: 1-550 <LR>  
A:Cross-references: GB:M74889  
A>Note: the nucleotide sequence is not complete  
A>Note: translation of the nucleotide sequence is not complete  
C:Superfamily: hepatitis C virus genome polyprotein  
C:Keywords: envelope protein; glycoprotein; nonstructural protein; nucleocapsid; polyprotein  
F:1-190/Product: nucleocapsid protein C [status predicted <CPC>  
F:191-380/Product: envelope protein E1 [status predicted <EPI>  
F:381-514/Product: envelope protein E2 or nonstructural protein NS1 (fragment) [status predicted <NS1>  
F:515-550/Product: nonstructural protein NS5 (fragment) [status predicted <NS5>  
F:196-233,250,305,416,422,429,447/Binding site: carbohydrate (Asn) (covalent) [status predicted <ASP>  
F:196-233,250,305,416,422,429,447/Binding site: carbohydrate (Asn) (covalent) [status predicted <ASP>  
Alignment Scores:  
Pred. No.: 5,28e-88 Length: 550  
Score: 1045.00 Matches: 188  
Percent Similarity: 95.19% Conservative: 10  
Best Local Similarity: 90.38% Mismatches: 10  
Query Match: 90.48% Indels: 0  
Gaps: 0  
DB: 2  
US-09-899-303a-7 (1-633) x JH0711 (1-550)  
QY 4 TTGGGTAAAGTCATGATACCTTACGCTTCGGCTTCGCCGACCTCATGGGGTACATTCCG 63  
Db 119 LeuGlyValValIleAspThrLeuThrCysGlyPheAlaAspLeuMetGlyTyrIlePro 138  
QY 64 CTCGTGGCGCCCCCTTAGGGGGTCTGCCAGAGCCCTGGGCATGGGGTCCGGTTCTG 123  
Db 139 LeuValGlyAlaProIleuGlyGlyAlaValaIleuAlaIleuAlaIleuAlaIleu 158  
QY 124 GAAGACGGCGTGAACATGACAGAGGAATTTGGCTGGTCTCTCTTCTATCTCTCTC 183  
Db 159 GluAspGlyValaIleuThrGlyAlaIleuProGlyCysSerPheSerIlePheIleu 178  
QY 184 TTGGCTTACTGCTCTGTCAGCATTCACATTCCTCCGTTATGAGAGTCCGACGTTGC 243  
Db 179 LeuAlaLeuLeuSerCysLeuThrThrProAlaSerAlaTyrGlyValaIleuAlaSer 198  
QY 244 GGGATGACACATGTCAGCAAGACGACTGCTCCAACTCAAGCATGTGTATGAGAGCGAC 303  
Db 199 GlyValIleThrIleValIleThrAsnAspCysSerAsnThrAsnIleValIleGlyThrAlaAsp 218  
QY 304 ATGATCATGCACACCCCGGGTGGCTGCGCTGCTCGGAGAACACTTCCCGCTGC 363  
Db 219 MetIleMetHisThrProGlyCysAlaIleProCysValaIleArgAspAsnGlyThrSerArgCys 238  
QY 364 TGGGTAGCGCTACCCCGACGCTGAGAGTGAAGAGCGACGCTCCCATACGACAATA 423  
Db 239 TrpValAlaLeuThrProThrLeuAlaIleAlaArgAsnAlaSerIleProThrThrAlaIle 258  
QY 424 CGAGCGCACGTCGATTTGCTGCTGTCGGGGGCTGCTTCTTCCGCTATAGCTGGGG 483  
Db 259 ArgArgHisIleIleAspLeuValGlyAlaIleAlaThrPheCysSerAlaMetGlyValGly 278  
QY 484 GATCTGCGGATGTCTTCTCTGCTGCTCCAGCTGTTCACATCTGCTCCGCGGAT 543  
Db 279 AspLeuCysGlySerIlePheLeuValSerIleIleuThrPheSerProArgGlnHis 298  
QY 544 GAGACGGTGCAGACTGCAATTTGCTCAATATCCCGGACATACAGTACAGTACGATG 603  
Db 299 GluThrValGlnAspCysAsnGlySerIleuThrProGlyHisValThrGlyHisArgMet 318  
QY 604 GCTTGGATATGATGATGAAGTGG 627  
Db 319 AlaTrpAspMetMetMetAsnTrp 326  
RESULT 15  
J01926  
polyprotein - hepatitis C virus (isolate HCV-476)  
N:Contains: C protein; E1 protein; E2/NS1 protein

C:Species: hepatitis C virus  
C:Date: 03-May-1994 #sequence,revision 03-May-1994 #text,change 17-Nov-2000  
C:Accession: J01926  
R:Abbe, K.; Inchausti, G.; Fujisawa, K.  
J. Gen. Virol. 73, 2725-2729, 1992  
A:Title: Genomic characterization and mutation rate of hepatitis C virus isolated from  
A:Reference number: J01925; MUID:93019030; PMID:1383400  
A:Accession: J01926  
A:Molecule type: mRNA  
A:Residues: 1-523 <ABE>  
A:Cross-references: DDBJ:D10687  
C:Superfamily: hepatitis C virus genome polyprotein  
C:Keywords: polyprotein  
F:1-191/Product: C protein [status predicted <CPR>  
F:192-383/Product: E1 protein [status predicted <E1R>  
F:384-523/Product: E2/NS1 protein [status predicted <E2P>  
Alignment Scores:  
Pred. No.: 6,52e-88 Length: 523  
Score: 1044.00 Matches: 192  
Percent Similarity: 96.15% Conservative: 8  
Best Local Similarity: 92.31% Mismatches: 8  
Query Match: 90.39% Indels: 0  
Gaps: 0  
DB: 2  
US-09-899-303a-7 (1-633) x J01926 (1-523)  
QY 4 TTGGGTAAAGTCATGATACCTTACGCTTCGGCTTCGCCGACCTCATGGGGTACATTCCG 63  
Db 119 LeuGlyValValIleAspThrLeuThrCysGlyPheAlaAspLeuMetGlyTyrIlePro 138  
QY 64 CTCGTGGCGCCCCCTTAGGGGGTCTGCCAGAGCCCTGGGCATGGGGTCCGGTTCTG 123  
Db 139 LeuValGlyAlaProIleuGlyGlyAlaValaIleuAlaIleuAlaIleuAlaIleu 158  
QY 124 GAAGACGGCGTGAACATGACAGAGGAATTTGGCTGGTCTCTCTTCTATCTCTCTC 183  
Db 159 GluAspGlyValaIleuThrGlyAlaIleuProGlyCysSerPheSerIlePheIleu 178  
QY 184 TTGGCTTACTGCTCTGTCAGCATTCACATTCCTCCGTTATGAGAGTCCGACGTTGC 243  
Db 179 SerAlaLeuMetSerCysLeuThrThrProAlaSerAlaTyrGlyValaIleuAlaSer 198  
QY 244 GGGATGACACATGTCAGCAAGACGACTGCTCCAACTCAAGCATGTGTATGAGAGCGAC 303  
Db 199 GlyIleThrIleValIleThrAsnAspCysSerAsnThrIleAlaTyrGlyAlaIleGly 218  
QY 304 ATGATCATGCACACCCCGGGTGGCTGCGCTGCTCGGAGAACACTTCCCGCTGC 363  
Db 219 MetIleMetHisThrProGlyCysValaIleProCysValaIleArgAspAsnSerSerArgCys 238  
QY 364 TGGGTAGCGCTACCCCGACGCTGAGAGTGAAGAGCGACGCTCCCATACGACAATA 423  
Db 239 TrpValAlaLeuThrProThrLeuAlaIleAlaArgAsnAlaSerIleProThrThrAlaIle 258  
QY 424 CGAGCGCACGTCGATTTGCTGCTGTCGGGGGCTGCTTCTTCCGCTATAGCTGGGG 483  
Db 259 ArgArgHisIleIleAspLeuValGlyAlaIleAlaThrPheCysSerAlaMetGlyValGly 278  
QY 484 GATCTGCGGATGTCTTCTCTGCTGCTCCAGCTGTTCACATCTGCTCCGCGGAT 543  
Db 279 AspLeuCysGlySerValPheLeuValSerIleIleuThrPheSerProArgGlnHis 298  
QY 544 GAGACGGTGCAGACTGCAATTTGCTCAATATCCCGGACATACAGTACAGTACGATG 603  
Db 299 GluThrValGlnAspCysAsnGlySerIleuThrProGlyHisValSerGlyHisArgMet 318  
QY 604 GCTTGGATATGATGATGAAGTGG 627  
Db 319 AlaTrpAspMetMetMetAsnTrp 326  
Search completed: February 12, 2003, 08:57:19

Job time : 36.5 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.3  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - protein search, using frame\_plus\_n2p model

Run on: February 12, 2003, 08:55:54 ; Search time 241 Seconds

(withn alignments)  
3385.855 Million cell updates/sec

Title: US-09-899-303A-7

Perfect score: 1155

Sequence: 1 ATGTTGGTGAAGTCATCGA.....TGATGATGACGTGTAATAG 633

Scoring table:

BL0SUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 4569144 seqs, 644733110 residues

Total number of hits satisfying chosen parameters: 9138288

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODE=frame-n2p.model -DEV=nlp  
-O/cgn2.1/USPTO.spool/US09899303.r/unat.12022003.085450.28326/app.query.fasta.1.775  
-DB=Pending\_Patents\_AA\_Main -QPM=fastan -SUFFIX=n2p.rapm -MINMATCH=0.1  
-LOOPEL=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=bl0sum62  
-TRANS=human40.ccl -LIST=45 -DOCALIGN=200 -THR=SCORE=pcr -THR\_MAX=100  
-THR\_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0  
-MAXLEN=2000000000 -USER=US09899303.@CGN.1.1.175.@unat.12022003.085450.28326  
-NCPU=6 -ICPU=3 -NO\_XLPRY -NO\_MMAP -LARGEQUERY -NEG\_SCORES=0 -WAIT -LONGLOG  
-DEV\_TIMEOUT=120 -WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

1: Pending\_Patents\_AA\_Main:\*  
2: /cgn2.6/ptodata/1/paa/PCTUS.COMB.pep:\*  
3: /cgn2.6/ptodata/1/paa/US07.COMB.pep:\*  
4: /cgn2.6/ptodata/1/paa/US08.COMB.pep:\*  
5: /cgn2.6/ptodata/1/paa/US081.COMB.pep:\*  
6: /cgn2.6/ptodata/1/paa/US082.COMB.pep:\*  
7: /cgn2.6/ptodata/1/paa/US083.COMB.pep:\*  
8: /cgn2.6/ptodata/1/paa/US084.COMB.pep:\*  
9: /cgn2.6/ptodata/1/paa/US085.COMB.pep:\*  
10: /cgn2.6/ptodata/1/paa/US086.COMB.pep:\*  
11: /cgn2.6/ptodata/1/paa/US087.COMB.pep:\*  
12: /cgn2.6/ptodata/1/paa/US088.COMB.pep:\*  
13: /cgn2.6/ptodata/1/paa/US089.COMB.pep:\*  
14: /cgn2.6/ptodata/1/paa/US090.COMB.pep:\*  
15: /cgn2.6/ptodata/1/paa/US091.COMB.pep:\*  
16: /cgn2.6/ptodata/1/paa/US092.COMB.pep:\*  
17: /cgn2.6/ptodata/1/paa/US093.COMB.pep:\*  
18: /cgn2.6/ptodata/1/paa/US094.COMB.pep:\*  
19: /cgn2.6/ptodata/1/paa/US095.COMB.pep:\*  
20: /cgn2.6/ptodata/1/paa/US096.COMB.pep:\*  
21: /cgn2.6/ptodata/1/paa/US097.COMB.pep:\*  
22: /cgn2.6/ptodata/1/paa/US098.COMB.pep:\*  
23: /cgn2.6/ptodata/1/paa/US099.COMB.pep:\*  
24: /cgn2.6/ptodata/1/paa/US100.COMB.pep:\*  
25: /cgn2.6/ptodata/1/paa/US101.COMB.pep:\*  
26: /cgn2.6/ptodata/1/paa/US102.COMB.pep:\*  
27: /cgn2.6/ptodata/1/paa/US60.COMB.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	1118	96.8	209	22	US-09-899-303-8	Sequence 8, Appl
2	1118	96.8	209	22	US-09-899-303A-8	Sequence 8, Appl
3	1118	96.8	209	23	US-09-973-025-8	Sequence 8, Appl
4	1118	96.8	209	23	US-09-995-791-8	Sequence 8, Appl
5	1118	96.8	209	23	US-09-995-808-8	Sequence 8, Appl
6	1118	96.8	209	23	US-09-995-860-8	Sequence 8, Appl
7	1118	96.8	209	24	US-10-020-510-8	Sequence 8, Appl
8	1118	96.8	209	25	US-10-128-590-88	Sequence 88, Appl
9	1113	96.4	263	22	US-09-899-303-6	Sequence 6, Appl
10	1113	96.4	263	22	US-09-899-303A-6	Sequence 6, Appl
11	1113	96.4	263	23	US-09-973-025-6	Sequence 6, Appl
12	1113	96.4	263	23	US-09-995-791-6	Sequence 6, Appl
13	1113	96.4	263	23	US-09-995-808-6	Sequence 6, Appl
14	1113	96.4	263	23	US-09-995-860-6	Sequence 6, Appl
15	1113	96.4	263	24	US-10-020-510-6	Sequence 6, Appl
16	1109	96.0	210	22	US-09-899-303-14	Sequence 6, Appl
17	1109	96.0	210	22	US-09-899-303A-14	Sequence 14, Appl
18	1109	96.0	210	23	US-09-973-025-14	Sequence 14, Appl
19	1109	96.0	210	23	US-09-995-791-14	Sequence 14, Appl
20	1109	96.0	210	23	US-09-995-808-14	Sequence 14, Appl
21	1109	96.0	210	23	US-09-995-860-14	Sequence 14, Appl
22	1109	96.0	210	24	US-10-020-510-14	Sequence 14, Appl
23	1108	95.9	692	22	US-09-899-303-48	Sequence 48, Appl
24	1108	95.9	692	22	US-09-899-303A-48	Sequence 48, Appl
25	1108	95.9	692	23	US-09-973-025-48	Sequence 48, Appl
26	1108	95.9	692	23	US-09-995-791-48	Sequence 48, Appl
27	1108	95.9	692	23	US-09-995-808-48	Sequence 48, Appl
28	1108	95.9	692	23	US-09-995-860-48	Sequence 48, Appl
29	1108	95.9	692	24	US-10-020-510-48	Sequence 48, Appl
30	1108	95.9	809	22	US-09-899-303-50	Sequence 50, Appl
31	1108	95.9	809	22	US-09-899-303A-50	Sequence 50, Appl
32	1108	95.9	809	23	US-09-973-025-50	Sequence 50, Appl
33	1108	95.9	809	23	US-09-995-791-50	Sequence 50, Appl
34	1108	95.9	809	23	US-09-995-808-50	Sequence 50, Appl
35	1108	95.9	809	23	US-09-995-860-50	Sequence 50, Appl
36	1108	95.9	809	24	US-10-020-510-50	Sequence 50, Appl
37	1094	94.7	450	10	US-08-635-886B-188	Sequence 188, App
38	1094	94.7	450	10	US-08-635-886C-188	Sequence 188, App
39	1094	94.7	513	3	US-07-790-382A-31	Sequence 31, Appl
40	1094	94.7	733	3	US-07-866-045-12	Sequence 12, Appl
41	1092	94.5	3010	19	US-09-539-601-3	Sequence 3, Appl
42	1092	94.5	3010	19	US-09-539-601-21	Sequence 21, Appl
43	1092	94.5	3010	19	US-09-539-601-27	Sequence 27, Appl
44	1092	94.5	3010	19	US-09-539-601-33	Sequence 33, Appl
45	1081	93.6	3010	20	US-09-662-454-3	Sequence 3, Appl

#### ALIGNMENTS

RESULT 1  
US-09-899-303-8  
Sequence 8, Application US/09899303  
GENERAL INFORMATION:  
APPLICANT: MAERTENS, GEERT  
BOSMAN, FONS  
DE MARTYNOFF, GUY  
BUYSE, MARIE-ANGE  
TITLE OF INVENTION: PURIFIED HEPATITIS C VIRUS ENVELOPE  
PROTEINS FOR DIAGNOSTIC AND THERAPEUTIC USE  
NUMBER OF SEQUENCES: 111  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: NIXON & VANDERHYTE P.C.  
STREET: 1100 NORTH GLEBE ROAD  
CITY: ARLINGTON

STATE: VIRGINIA  
COUNTRY: U.S.A.  
ZIP: 22201-4714  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.25 (EPO)  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/899,303  
FILING DATE: 06-Jul-2001  
CLASSIFICATION: <Unknown>  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/612,973  
FILING DATE: 11-MAR-1996  
ATTORNEY/AGENT INFORMATION:  
NAME: BYRNE, THOMAS E.  
REGISTRATION NUMBER: 32,205  
REFERENCE/DOCKET NUMBER: 1487-10  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (703) 816-4000  
TELEFAX: (703) 816-4100  
INFORMATION FOR SEQ ID NO: 8:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 209 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
SEQUENCE DESCRIPTION: SEQ ID NO: 8:  
US-09-899-303-8  
Alignment Scores:  
Pred. No.: 5,77e-96 Length: 209  
Score: 1118.00 Matches: 209  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 96.80% Indels: 0  
Gaps: 0  
DB: 22  
US-09-899-303A-7 (1-633) x US-09-899-303-8 (1-209)  
QY 1 ATGTTGGGTAAGGTCATCATACCTTACGTGCGGCTTCCGACCTCATGGGTCATATT 60  
DB 1 MetLeuGlyLysValIleAspThrLeuThrCysGlyPheAlaAspLeuMetGlyTyrIle 20  
QY 61 CCGCTGTCGGGCGCCCGCTAGGGGGTGTGCCAGAGCCCTGGCGCATGGCTCCGGGTT 120  
DB 21 ProlLeuValGlyAlaProlLeuGlyGlyAlaAlaArgAlaLeuAlaHisGlyValArgVal 40  
QY 121 CTGGAGAGGGGGTGAATGCAACAGGGAATTTGCTGGTCTTCTTCTATCTTC 180  
DB 41 LeuGlnAspGlyValAsnGlyAlaThrGlyAsnLeuProlGlyCysSerPheSerIlePhe 60  
QY 181 CTTTGGCTTTTACTGCTGTGTGACCATTCAGCTTCCGCTTATGAGGCGCAAGTG 240  
DB 61 LeuLeuAlaLeuLeuSerGlyLeuThrIleProlAspSerIleValGlyValArgVal 80  
QY 241 TCCGGGATGTACATGTCAAGAGACTGCTCCAACTCAAGCATTTGTATGAGGAGCG 300  
DB 81 SerGlyMetTyrHisValIleHisAspCysSerAsnSerSerIleValGlyGluAlaAla 100  
QY 301 GACATGATATGACACACCCCGGGTCCGCTGCGCTTCCGAGACAACTTCTCCCGC 360  
DB 101 AspMetIleMetHisThrProlGlyCysValProCysValArgGlyAsnAsnSerSerArg 120  
QY 361 TCGTGGGTAGCGCTCAACCCGAGCTCCAGCTAGGAGCGCAGGCTCCCACTAGAGCA 420  
DB 121 CysTrpValAlaLeuThrThrThrIleuAlaAlaAlaGlnAlaSerValProlThrThrThr 140  
QY 421 ATAGACAGCCACGATGATTTGCTGTGGGCGGCTCTTCTGTTCCGCTATGACGTG 480  
DB 141 IleArgGlnHisValAspLeuLeuValGlyAlaAlaAlaPheCysSerAlaMetGlyVal 160

QY 481 GGGGATCTCGCGGATCTCTTCTCTGTCCTCCAGCTGTACACATCTGCGCTCCCGG 540  
DB 161 GlyAspLeuGlySerValPheLeuValSerGlnLeuPheThrIleSerProlArgArg 180  
QY 541 CATGACAGCGGTGACGACATGCTCATCTATCCGCGCCACATACAGTCAACCGT 600  
DB 181 HisGlnThrValIleAspCysAsnGlySerIleTyrProlHisIleThrGlyHisArg 200  
QY 601 ATGGCTTGGGATGATGATGATGATG 627  
DB 201 MetAlaTrpAspMetMetMetAsnTrp 209  
RESULT 2  
US-09-899-303A-8  
Sequence 8, Application US/09899303A  
GENERAL INFORMATION:  
APPLICANT: MAERTENS, GEERT  
BOSMAN, PONS  
DE MARTYNOFF, GUY  
BOYSE, MARIE-ANGE  
TITLE OF INVENTION: PURIFIED HEPATITIS C VIRUS ENVELOPE  
PROTEINS FOR DIAGNOSTIC AND THERAPEUTIC USE  
NUMBER OF SEQUENCES: 111  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: NIXON & VANDERHAYE P.C.  
STREET: 1100 NORTH GLEBE ROAD  
CITY: ARLINGTON  
STATE: VIRGINIA  
COUNTRY: U.S.A.  
ZIP: 22201-4714  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.25 (EPO)  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/899,303A  
FILING DATE: 06-Jul-2001  
CLASSIFICATION: <Unknown>  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US/08/612,973  
FILING DATE: 11-MAR-1996  
ATTORNEY/AGENT INFORMATION:  
NAME: BYRNE, THOMAS E.  
REGISTRATION NUMBER: 32,205  
REFERENCE/DOCKET NUMBER: 1487-10  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (703) 816-4000  
TELEFAX: (703) 816-4100  
SEQUENCE CHARACTERISTICS:  
LENGTH: 209 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
SEQUENCE DESCRIPTION: SEQ ID NO: 8:  
US-09-899-303A-8  
Alignment Scores:  
Pred. No.: 5,77e-96 Length: 209  
Score: 1118.00 Matches: 209  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 96.80% Indels: 0  
Gaps: 0  
DB: 22  
US-09-899-303A-7 (1-633) x US-09-899-303A-8 (1-209)  
QY 1 ATGTTGGGTAAGGTCATCATACCTTACGTGCGGCTTCCGACCTCATGGGTCATATT 60  
DB 1 MetLeuGlyLysValIleAspThrLeuThrCysGlyPheAlaAspLeuMetGlyTyrIle 20  
QY 61 CCGCTGTCGGGCGCCCGCTAGGGGGTGTGCCAGAGCCCTGGCGCATGGCTCCGGGTT 120



```
FILE REFERENCE: 2551-68
CURRENT APPLICATION NUMBER: US/09/995,791
CURRENT FILING DATE: 2001-11-29
NUMBER OF SEQ ID NOS: 122
SOFTWARE: PatentIn 3.1
SEQ ID NO 8
LENGTH: 209
TYPE: PRF
ORGANISM: Hepatitis C virus
US-09-995-791-8

Alignment Scores:
Pred. No.: 5 77e-96 Length: 209
Score: 1118.00 Matches: 209
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 96.80% Indels: 0
DB: 23 Gaps: 0

US-09-899-303A-7 (1-633) x US-09-995-791-8 (1-209)

QY 1 ATGTGGGTAAAGTCAATCGATACCTTACGTGGGCTTCCGACCTCATGGGGTACATT 60
DB 1 MetLeuGlyLysValIleAspThrLeuThrCysGlyPheAlaAspLeuMetGlyTyrIle 20
QY 61 CCGCTGTCGGGCCCCCTAGGGGGTGTGCCAGAGCCCTGGCGCATGGCGTCCGGGTT 120
DB 21 ProLeuValGlyAlaProLeuGlyGlyAlaAlaArgAlaLeuAlaHisGlyValArgVal 40
QY 121 CTGGAAGAGCGGGTGAACATGCAACAGGAATTTCCTGGTTGCTTCTCTATCTTC 180
DB 41 LeuGluAspGlyValAsnTyrAlaThrGlyAsnLeuProGlyCysSerPheSerIlePhe 60
QY 181 CTCCTGGCTTACTGCTGCTGTGACCAATTCAGCTTCGCTTATGAGTGGCAACGTG 240
DB 61 LeuLeuAlaLeuLeuSerCysLeuThrIleProAlaSerAlaTyrGluValArgAsnVal 80
QY 241 TCCGGGATACCATGTCAAGAACGACTGCTCCAACTCAAGCATTTGTATAGGACGG 300
DB 81 SerGlyMetTyrHisValIleThrAsnAspCysSerAsnSerSerIleValTyrGluAlaIle 100
QY 301 GACATGATCATGCACACCCCGGGGTGCGCCCTGGCTTCGGGAGAACACTCTTCCCG 360
DB 101 AspMetIleMetHisThrProGlyCysValProCysValArgGluAsnAsnSerSerArg 120
QY 361 TCGTGGGTAGCGCTCACCCCAAGCTCGAGCTAGGAAAGCCAGCGTCCCACTAGACA 420
DB 121 CysTrpValAlaLeuThrProThrLeuAlaAlaArgAlaAsnAlaSerValProThrThrThr 140
QY 421 ATACGACGCCAGCTGCATTTGCTGCTGGGGGGGCTGCTTTCTGTTCCGCTATGACGTG 480
DB 141 IleArgArgHisValAspLeuLeuValGlyAlaAlaAlaPheCysSerAlaMetTyrVal 160
QY 481 GGGGATCTGGGGATCTGCTCTTCCTGCTCCAGCTTTCACCACTTCGCGCGCGG 540
DB 161 GlyAspLeuCysGlySerValPheLeuValSerGlnLeuPheThrIleSerProArgArg 180
QY 541 CATGAGACGGTCCAGACTGCAATTGCTCAATCTATCCGGCCACATAACAGTCAACGT 600
DB 181 HisGluThrValGlnAspCysAsnSerIleTyrProGlyHisIleThrGlyHisArg 200
QY 601 ATGGCTTGGGATATGATGATGAACCTGG 627
DB 201 MetaLatrpAspMetMetMetAsnTrp 209

RESULT 5
US-09-995-808-8
; Sequence 8, Application US/09995808
; GENERAL INFORMATION:
; APPLICANT: Innogenetics N.V.
; TITLE OF INVENTION: Purified hepatitis C virus envelope proteins for diagnostic and
; FILE REFERENCE: 2551-70
```

```
CURRENT APPLICATION NUMBER: US/09/995,808
CURRENT FILING DATE: 2001-11-29
NUMBER OF SEQ ID NOS: 122
SOFTWARE: PatentIn 3.1
SEQ ID NO 8
LENGTH: 209
TYPE: PRF
ORGANISM: Hepatitis C virus
US-09-995-808-8

Alignment Scores:
Pred. No.: 5 77e-96 Length: 209
Score: 1118.00 Matches: 209
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 96.80% Indels: 0
DB: 23 Gaps: 0

US-09-899-303A-7 (1-633) x US-09-995-808-8 (1-209)

QY 1 ATGTGGGTAAAGTCAATCGATACCTTACGTGGGCTTCCGACCTCATGGGGTACATT 60
DB 1 MetLeuGlyLysValIleAspThrLeuThrCysGlyPheAlaAspLeuMetGlyTyrIle 20
QY 61 CCGCTGTCGGGCCCCCTAGGGGGTGTGCCAGAGCCCTGGCGCATGGCGTCCGGGTT 120
DB 21 ProLeuValGlyAlaProLeuGlyGlyAlaAlaArgAlaLeuAlaHisGlyValArgVal 40
QY 121 CTGGAAGAGCGGGTGAACATGCAACAGGAATTTCCTGGTTGCTTCTCTATCTTC 180
DB 41 LeuGluAspGlyValAsnTyrAlaThrGlyAsnLeuProGlyCysSerPheSerIlePhe 60
QY 181 CTCCTGGCTTACTGCTGCTGTGACCAATTCAGCTTCGCTTATGAGTGGCAACGTG 240
DB 61 LeuLeuAlaLeuLeuSerCysLeuThrIleProAlaSerAlaTyrGluValArgAsnVal 80
QY 241 TCCGGGATACCATGTCAAGAACGACTGCTCCAACTCAAGCATTTGTATAGGACGG 300
DB 81 SerGlyMetTyrHisValIleThrAsnAspCysSerAsnSerSerIleValTyrGluAlaIle 100
QY 301 GACATGATCATGCACACCCCGGGGTGCGCCCTGGCTTCGGGAGAACACTCTTCCCG 360
DB 101 AspMetIleMetHisThrProGlyCysValProCysValArgGluAsnAsnSerSerArg 120
QY 361 TCGTGGGTAGCGCTCACCCCAAGCTCGAGCTAGGAAAGCCAGCGTCCCACTAGACA 420
DB 121 CysTrpValAlaLeuThrProThrLeuAlaAlaArgAlaAsnAlaSerValProThrThrThr 140
QY 421 ATACGACGCCAGCTGCATTTGCTGCTGGGGGGGCTGCTTTCTGTTCCGCTATGACGTG 480
DB 141 IleArgArgHisValAspLeuLeuValGlyAlaAlaAlaPheCysSerAlaMetTyrVal 160
QY 481 GGGGATCTGGGGATCTGCTCTTCCTGCTCCAGCTTTCACCACTTCGCGCGCGG 540
DB 161 GlyAspLeuCysGlySerValPheLeuValSerGlnLeuPheThrIleSerProArgArg 180
QY 541 CATGAGACGGTCCAGACTGCAATTGCTCAATCTATCCGGCCACATAACAGTCAACGT 600
DB 181 HisGluThrValGlnAspCysAsnSerIleTyrProGlyHisIleThrGlyHisArg 200
QY 601 ATGGCTTGGGATATGATGATGAACCTGG 627
DB 201 MetaLatrpAspMetMetMetAsnTrp 209

RESULT 6
US-09-995-860-8
; Sequence 8, Application US/09995860
; GENERAL INFORMATION:
; APPLICANT: Innogenetics N.V.
; TITLE OF INVENTION: Purified hepatitis C virus envelope proteins for diagnostic an
; FILE REFERENCE: 2551-69
; CURRENT APPLICATION NUMBER: US/09/995,860
```

;; CURRENT FILING DATE: 2001-11-29  
;; NUMBER OF SEQ ID NOS: 122  
;; SOFTWARE: PatentIn 3.1  
;; SEQ ID NO 8  
;; LENGTH: 209  
;; TYPE: PRF  
;; ORGANISM: Hepatitis C virus  
US-09-995-860-8

## Alignment Scores:

Pred. No.:	5, 77e-96	length:	209
Score:	1118.00	Matches:	209
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	96.80%	Indels:	0
DB:	23	Gaps:	0

US-09-899-303A-7 (1-633) x US-09-995-860-8 (1-209)

QY 1 ATGTTGGGTAAAGTTCATCGATACCTTACGTGCGGCTTCGCCGACCTCATGGGGTACATT 60  
|||||  
Db 1 MetLeuGlYysValIleAspThrLeuThrCysGlYpheaIlaSpLeuMetGlYTYrIle 20  
QY 61 CCGCTGCTGGCGCCGCCCTAGGGGGTGTGCCAGAGACCTGGCGCATGGCGTCCGGGTT 120  
|||||  
Db 21 ProLeuValGlYAlaProLeuGlYGlYAlaAlaArgAlaLeuAlaHisGlYValaArgVal 40  
QY 121 CTGAAGACGGCGGTGAACATTCAGCAACAGGAATTTGGCTGTTGCTCTTCTCATCTTC 180  
|||||  
Db 41 LeuGlUnspolYAlaAntYrAlaThrGlYAsnLeuProGlYsSerPheSerIlePhe 60  
QY 181 CTCTTGCTTTACTGTCTGCTGTGACCATTCAGCTTCCGCTTATGAGTGCGCAACGTG 240  
|||||  
Db 61 LeuLeuAlaLeuLeuSerCysLeuThrIleProAlaSerAlaTyrGlUValaArgAsnVal 80  
QY 241 TCCGGGATGACCATGTCACGAAGAGACTGCTCCAACTCAAGCATTTGTATGAGCGACGC 300  
|||||  
Db 81 SerGlYMetYrHisValThrAsnAspCysSerAsnSerSerIleValTYrGlUAlaAla 100  
QY 301 GACATGATCATGACACCCCGGGGTGGCTGCCCTGGCTTGGGAGAACATCTTCCCGC 360  
|||||  
Db 101 AspMetIleMetHisThrProGlYCysValProCysValaArgGlUnsnAsnSerSerArg 120  
QY 361 TGCTGGGTAGCGCTCAACCCCGCTCGACGCTGAGTAAGAACGCCAGCTCCCACTAGCACA 420  
|||||  
Db 121 CysTrpValAlaLeuThrProThrLeuAlaAlaArgAsnAlaSerValProThrThrThr 140  
QY 421 ATACGACGCCACGTCGATTGCTGCTGGGGCGGCTGCTTCTTGTTCGGTATGACGTG 480  
|||||  
Db 141 IleArgArGHisValAspLeuLeuValGlYAlaAlaAlaPheCysSerAlaMetTYrVal 160  
QY 481 GGGGATCTGCGGATCTGCTCTCCGCTCCAGCTGTTCACCATCTCGCGTCCGCGG 540  
|||||  
Db 161 GlYAspLeuCysGlYSerValaPheLeuValaSerGlnLeuPheThrIleSerProAlaArg 180  
QY 541 CATGACAGCGGTGACGAGACTGCAATTGCTCAATCTATCCCGGCACATMACAGGTACCGT 600  
|||||  
Db 181 HisGlUThrValaGlnAspCysAsnCysSerIleTYrProGlYHisIleThrGlYHisArg 200  
QY 601 ATGGCTTGGATATGATGATGAACGTG 627  
|||||  
Db 201 MetAlaTrpAspMetMetMetAsnTrp 209

RESULT 7  
US-10-020-510-8

;; Sequence 8, Application US/10020510  
;; GENERAL INFORMATION:  
;; APPLICANT: Immunogenetics N.V.  
;; TITLE OF INVENTION: Purified hepatitis C virus envelope proteins for diagnostic and  
;; FILE REFERENCE: 2551-72  
;; CURRENT FILING DATE: 2002-07-15  
;; CURRENT APPLICATION NUMBER: US/10/020, 510

;; NUMBER OF SEQ ID NOS: 122  
;; SOFTWARE: PatentIn 3.1  
;; SEQ ID NO 8  
;; LENGTH: 209  
;; TYPE: PRF  
;; ORGANISM: Hepatitis C virus  
US-10-020-510-8

## Alignment Scores:

Pred. No.:	5, 77e-96	length:	209
Score:	1118.00	Matches:	209
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	96.80%	Indels:	0
DB:	24	Gaps:	0

US-09-899-303A-7 (1-633) x US-10-020-510-8 (1-209)

QY 1 ATGTTGGGTAAAGTTCATCGATACCTTACGTGCGGCTTCGCCGACCTCATGGGGTACATT 60  
|||||  
Db 1 MetLeuGlYysValIleAspThrLeuThrCysGlYpheaIlaSpLeuMetGlYTYrIle 20  
QY 61 CCGCTGCTGGCGCCGCCCTAGGGGGTGTGCCAGAGACCTGGCGCATGGCGTCCGGGTT 120  
|||||  
Db 21 ProLeuValGlYAlaProLeuGlYGlYAlaAlaArgAlaLeuAlaHisGlYValaArgVal 40  
QY 121 CTGAAGACGGCGGTGAACATTCAGCAACAGGAATTTGGCTGTTGCTCTTCTCATCTTC 180  
|||||  
Db 41 LeuGlUnspolYAlaAntYrAlaThrGlYAsnLeuProGlYsSerPheSerIlePhe 60  
QY 181 CTCTTGCTTTACTGTCTGCTGTGACCATTCAGCTTCCGCTTATGAGTGCGCAACGTG 240  
|||||  
Db 61 LeuLeuAlaLeuLeuSerCysLeuThrIleProAlaSerAlaTyrGlUValaArgAsnVal 80  
QY 241 TCCGGGATGACCATGTCACGAAGAGACTGCTCCAACTCAAGCATTTGTATGAGCGACGC 300  
|||||  
Db 81 SerGlYMetYrHisValThrAsnAspCysSerAsnSerSerIleValTYrGlUAlaAla 100  
QY 301 GACATGATCATGACACCCCGGGGTGGCTGCCCTGGCTTGGGAGAACATCTTCCCGC 360  
|||||  
Db 101 AspMetIleMetHisThrProGlYCysValProCysValaArgGlUnsnAsnSerSerArg 120  
QY 361 TGCTGGGTAGCGCTCAACCCCGCTCGACGCTGAGTAAGAACGCCAGCTCCCACTAGCACA 420  
|||||  
Db 121 CysTrpValAlaLeuThrProThrLeuAlaAlaArgAsnAlaSerValProThrThrThr 140  
QY 421 ATACGACGCCACGTCGATTGCTGCTGGGGCGGCTGCTTCTTGTTCGGTATGACGTG 480  
|||||  
Db 141 IleArgArGHisValAspLeuLeuValGlYAlaAlaAlaPheCysSerAlaMetTYrVal 160  
QY 481 GGGGATCTGCGGATCTGCTCTCCGCTCCAGCTGTTCACCATCTCGCGTCCGCGG 540  
|||||  
Db 161 GlYAspLeuCysGlYSerValaPheLeuValaSerGlnLeuPheThrIleSerProAlaArg 180  
QY 541 CATGACAGCGGTGACGAGACTGCAATTGCTCAATCTATCCCGGCACATMACAGGTACCGT 600  
|||||  
Db 181 HisGlUThrValaGlnAspCysAsnCysSerIleTYrProGlYHisIleThrGlYHisArg 200  
QY 601 ATGGCTTGGATATGATGATGAACGTG 627  
|||||  
Db 201 MetAlaTrpAspMetMetMetAsnTrp 209

RESULT 8  
US-10-128-590-88

;; Sequence 88, Application US/10128590  
;; GENERAL INFORMATION:  
;; APPLICANT: Immunogenetics N.V.  
;; TITLE OF INVENTION: Core-glycosylated HCV envelope proteins  
;; FILE REFERENCE: 135 PCT  
;; CURRENT APPLICATION NUMBER: US/10/128, 590  
;; CURRENT FILING DATE: 2002-07-22  
;; NUMBER OF SEQ ID NOS: 98  
;; SOFTWARE: PatentIn version 3.1

SEQ ID NO 88  
 LENGTH: 209  
 TYPE: PRT  
 ORGANISM: hepatitis C virus  
 US-10-128-590-88

## Alignment Scores:

Pred. No.:	5.77e-96	Length:	209
Score:	1118.00	Matches:	209
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	96.80%	Indels:	0
DB:	25	Gaps:	0

US-09-899-303a-7 (1-633) x US-10-128-590-88 (1-209)

```

OY 1 ATGTTGGATGAGTATCATGATACCTTACGTGCGGCTTCCGACCTCATGGGGTACATT 60
    |||||
DB 1 MetLeuGlyLysValIleAspThrLeuThrCysGlyPheAlaAspLeuMetGlyTyrIle 20
OY 61 CCGCTCGTGGGCCCCCTAGGGGGTGTGCGACAGCCCTGGCGATGGCTCCGGGTT 120
    |||||
DB 21 ProLeuValGlyAlaProLeuGlyGlyAlaAlaArgAlaLeuAlaHisGlyValArgVal 40
OY 121 CTGGAAGAGGGGCTGAACATATGACACAGGAGATTGGCTGCTTCTCTATCTTC 180
    |||||
DB 41 LeuGluAspGlyValAsnTyrAlaThrGlyAsnLeuProGlyCysSerPheSerIlePhe 60
OY 181 CTCTGGCTTTACTGCTCTGCTGACACATTCAGCTTCGCTTATGAGTGGGACAGTG 240
    |||||
DB 61 LeuLeuAlaLeuLeuSerCysLeuThrIleProAlaSerAlaTyrGluValArgAsnVal 80
OY 241 TCCGGGATGTACCATGTCTACGACAGACTGTCCCAACTGATGTATGAGGACGG 300
    |||||
DB 81 SerGlyMetTyrHisValIleThrAsnAspCysSerAsnSerSerIleValTyrGluAlaAla 100
OY 301 GACATGATCATGACACACCCCGGGTCCGCTGGCTGGCGAGAGACACTCTCCCGCC 360
    |||||
DB 101 AspMetIleMetHisThrProGlyCysValProCysValArgGluAsnAsnSerSerArg 120
OY 361 TCGTGGGTAGCGCTACACCCCAAGCTCGACGTAGAGAGCCAGCGCTCCCACTACGACA 420
    |||||
DB 121 CysTrpValAlaLeuThrProThrIleuAlaAlaArgAsnAlaSerValProThrThrThr 140
OY 421 ATACAGACGCCAGCTGATTTGCTGCTGGGGGCGGCTTTCTGTTCCGCTATGACGTG 480
    |||||
DB 141 IleArgArgHisValAlaSerLeuValGlyAlaAlaAlaPheCysSerAlaMetTyrVal 160
OY 481 GGGGATCTCTGGGATCTGCTCTGCTCCGCTCCGACGTGTTCACCATCTCGCGCTGGCGG 540
    |||||
DB 161 GlyAspLeuCysGlySerValPheLeuValSerGlnLeuPheThrIleSerProAlaArg 180
OY 541 CATGAGACGGGTGACAGACTGCAATGCTCATCTCCGCGCACATACAGTCAAGTCAAGT 600
    |||||
DB 181 HisGluThrValGlnAspCysAsnSerIleTyrProGlyHisIleThrGlyHisArg 200
OY 601 ATGGCTGGGATATGATGATGATG 627
    |||||
DB 201 MetaIaTrpAspMetMetAsnTrp 209
  
```

## RESULT 9

US-09-899-303-6  
 Sequence 6, Application US/09899303  
 GENERAL INFORMATION:

APPLICANT: MAERTENS, GEERT

BOSMAN, FONS

DE MARTYNOFF, GUY

BOYSE, MARIE-ANGE

TITLE OF INVENTION: PURIFIED HEPATITIS C VIRUS ENVELOPE

PROTEINS FOR DIAGNOSTIC AND THERAPEUTIC USE

NUMBER OF SEQUENCES: 111

CORRESPONDENCE ADDRESSES:  
 ADDRESSEE: NIXON & VANDERHIVE P. C.

STREET: 1100 NORTH GLEBE ROAD  
 CITY: ARLINGTON  
 STATE: VIRGINIA  
 COUNTRY: U.S.A.  
 ZIP: 22201-4714

## COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/899,303

FILING DATE: 06-Jul-2001

CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/612,973

FILING DATE: 11-MAR-1996

ATTORNEY/AGENT INFORMATION:

NAME: BYRNE, THOMAS E.

REGISTRATION NUMBER: 32,205

REFERENCE/DOCKET NUMBER: 1487-10

TELECOMMUNICATION INFORMATION:

TELEPHONE: (703) 816-4000

TELEFAX: (703) 816-4100

INFORMATION FOR SEQ ID NO: 6:

SEQUENCE CHARACTERISTICS:

LENGTH: 263 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

SEQUENCE DESCRIPTION: SEQ ID NO: 6:

## US-09-899-303-6

US-09-899-303-6 (1-263)

## Alignment Scores:

Pred. No.:	1.79e-95	Length:	263
Score:	1113.00 <td>Matches:</td> <td>207</td>	Matches:	207
Percent Similarity:	100.00% <td>Conservative:</td> <td>2</td>	Conservative:	2
Best Local Similarity:	99.04% <td>Mismatches:</td> <td>0</td>	Mismatches:	0
Query Match:	96.36% <td>Indels:</td> <td>0</td>	Indels:	0
DB:	22	Gaps:	0

US-09-899-303a-7 (1-633) x US-09-899-303-6 (1-263)

```

OY 1 ATGTTGGATGAGTATCATGATACCTTACGTGCGGCTTCCGACCTCATGGGGTACATT 60
    |||||
DB 1 MetLeuGlyLysValIleAspThrLeuThrCysGlyPheAlaAspLeuValGlyTyrIle 20
OY 61 CCGCTCGTGGGCCCCCTAGGGGGTGTGCGACAGCCCTGGCGATGGCTCCGGGTT 120
    |||||
DB 21 ProLeuValGlyAlaProLeuGlyGlyAlaAlaArgAlaLeuAlaHisGlyValArgVal 40
OY 121 CTGGAAGAGGGGCTGAACATATGACACAGGAGATTGGCTGCTTCTCTATCTTC 180
    |||||
DB 41 LeuGluAspGlyValAsnTyrAlaThrGlyAsnLeuProGlyCysSerPheSerIlePhe 60
OY 181 CTCTGGCTTTACTGCTCTGCTGACACATTCAGCTTCGCTTATGAGTGGGACAGTG 240
    |||||
DB 61 LeuLeuAlaLeuLeuSerCysLeuThrValProAlaSerAlaTyrGluValArgAsnVal 80
OY 241 TCCGGGATGTACCATGTCTACGACAGACTGTCCCAACTCAAGATTGTATGAGGACGG 300
    |||||
DB 81 SerGlyMetTyrHisValIleThrAsnAspCysSerAsnSerSerIleValTyrGluAlaAla 100
OY 301 GACATGATCATGACACACCCCGGGTCCGCTGGCTGGCGAGAGACACTCTCCCGCC 360
    |||||
DB 101 AspMetIleMetHisThrProGlyCysValProCysValArgGluAsnAsnSerSerArg 120
OY 361 TCGTGGGTAGCGCTACACCCCAAGCTCGACGTAGAGAGCCAGCGCTCCCACTACGACA 420
    |||||
DB 121 CysTrpValAlaLeuThrProThrIleuAlaAlaArgAsnAlaSerValProThrThrThr 140
OY 421 ATACAGACGCCAGCTGATTTGCTGCTGGGGGCGGCTTTCTGTTCCGCTATGACGTG 480
    |||||
  
```





```
; TITLE OF INVENTION: therapeutic use.
; FILE REFERENCE: 2531-70
; CURRENT APPLICATION NUMBER: US/09/995,808
; CURRENT FILING DATE: 2001-11-29
; NUMBER OF SEQ ID NOS: 122
; SOFTWARE: PatentIn 3.1
; SEQ ID NO 6
; LENGTH: 263
; TYPE: PRT
; ORGANISM: Hepatitis C virus
US-09-995-808-6

Alignment Scores:
Pred. No.: 1.79e-95 Length: 263
Score: 1113.00 Matches: 207
Percent Similarity: 100.00% Conservative: 2
Best Local Similarity: 99.04% Mismatches: 0
Query Match: 96.36% Indels: 0
DB: 23 Gaps: 0

US-09-899-303a-7 (1-633) x US-09-995-808-6 (1-263)
QY 1 ATGTTGGGTAAAGTCATCGATACCTTACGTCGGGCTTCGCCGACCTCATGGGGTACATT 60
Db 1 MetLeuGlyysValIleAspThrLeuThrCysGlyPheAlaAspLeuValGlyTyrIle 20
QY 61 CCGCTCGTCGGCGCCCTAGAGGGGTGTCGCCAGAGCCCTGGCGCATGGCGTCCGGGTT 120
Db 21 ProLeuValGlyAlaProLeuGlyGlyAlaAlaArgAlaLeuAlaHisGlyValArgVal 40
QY 121 CTGGAGAGCGCGGTGAACATATGCAACAGGGAATTTGCCCTGGTTCCTTTCTATCTTC 180
Db 41 LeuGluAspGlyValAsnTyrAlaThrGlyAsnLeuProGlyCysSerPheSerIlePhe 60
QY 181 CTCCTGCTTACTGCTGCTGTCGACCATTCACACTCCGCTTATGAGGTGGCGAAGTG 240
Db 61 LeuLeuAlaLeuLeuSerCysLeuThrValProAlaSerAlaTyrGluValArgAsnVal 80
QY 241 TCCGGGATGTACCATGTACAGAACAGACTGCTCAACTCAAGCATTTGTATGAGGCAGCG 300
Db 81 SerGlyMetTyrHisValThrAsnAspCysSerAsnSerSerIleValTyrGluAlaIle 100
QY 301 GACATGATCATGCACACCCCGGGGTGCTGCTGCTGCGGGAACAACTCTTCCCGC 360
Db 101 AspMetIleMetHisThrProGlyCysValProCysValArgGluAsnAsnSerSerArg 120
QY 361 TGCCTGGGTAGCGGTACCCCGACGCTGCGAGCTGGAACGCGACGCTCCCACTAGACA 420
Db 121 CysThrPValAlaLeuThrProThrLeuAlaIleArgAsnAlaSerValProThrThrThr 140
QY 421 ATACGACGCCACGTGATTTGCTGCTGTTGGGGGCGGCTCTTTCCTGCTATGTACGTG 480
Db 141 IleArgThrHisValAspLeuLeuValGlyAlaAlaIlePheCysSerAlaMetTyrVal 160
QY 481 GGGGATCTTGGCGGATGTGCTTCTCTGCTGCCAGCTGTTCACCAATCTGCGCCGG 540
Db 161 G1yAspLeuCysGlySerValPheLeuValSerGlnLeuPheThrIleSerProArgArg 180
QY 541 CATGAGAGCGGTGACGAGCTGCAATTTGCTCAATCATCCGGGCCAATAAGCGTCACGCT 600
Db 181 HisGluThrValGlnAspCysAsnCysSerIleTyrProGlyHisIleThrGlyHisArg 200
QY 601 ATGCTTGGGATATGATGATGAACCTGG 627
Db 201 MetAlaThrPAspMetMetMetAlaThrP 209

RESULT 14
US-09-995-860-6
; Sequence 6, Application US/09995860
; GENERAL INFORMATION:
; APPLICANT: Innogenetics N.V.
; TITLE OF INVENTION: Purified hepatitis C virus envelope proteins for diagnostic and
; TITLE OF INVENTION: therapeutic use.
```

```
; FILE REFERENCE: 2531-69
; CURRENT APPLICATION NUMBER: US/09/995,860
; CURRENT FILING DATE: 2001-11-29
; NUMBER OF SEQ ID NOS: 122
; SOFTWARE: PatentIn 3.1
; SEQ ID NO 6
; LENGTH: 263
; TYPE: PRT
; ORGANISM: Hepatitis C virus
US-09-995-860-6

Alignment Scores:
Pred. No.: 1.79e-95 Length: 263
Score: 1113.00 Matches: 207
Percent Similarity: 100.00% Conservative: 2
Best Local Similarity: 99.04% Mismatches: 0
Query Match: 96.36% Indels: 0
DB: 23 Gaps: 0

US-09-899-303a-7 (1-633) x US-09-995-860-6 (1-263)
QY 1 ATGTTGGGTAAAGTCATCGATACCTTACGTCGGGCTTCGCCGACCTCATGGGGTACATT 60
Db 1 MetLeuGlyysValIleAspThrLeuThrCysGlyPheAlaAspLeuValGlyTyrIle 20
QY 61 CCGCTCGTCGGCGCCCTAGAGGGGTGTCGCCAGAGCCCTGGCGCATGGCGTCCGGGTT 120
Db 21 ProLeuValGlyAlaProLeuGlyGlyAlaAlaArgAlaLeuAlaHisGlyValArgVal 40
QY 121 CTGGAGAGCGCGGTGAACATATGCAACAGGGAATTTGCCCTGGTTCCTTTCTATCTTC 180
Db 41 LeuGluAspGlyValAsnTyrAlaThrGlyAsnLeuProGlyCysSerPheSerIlePhe 60
QY 181 CTCCTGCTTACTGCTGCTGTCGACCATTCACACTCCGCTTATGAGGTGGCGAAGTG 240
Db 61 LeuLeuAlaLeuLeuSerCysLeuThrValProAlaSerAlaTyrGluValArgAsnVal 80
QY 241 TCCGGGATGTACCATGTACAGAACAGACTGCTCAACTCAAGCATTTGTATGAGGCAGCG 300
Db 81 SerGlyMetTyrHisValThrAsnAspCysSerAsnSerSerIleValTyrGluAlaIle 100
QY 301 GACATGATCATGCACACCCCGGGGTGCTGCTGCTGCGGGAACAACTCTTCCCGC 360
Db 101 AspMetIleMetHisThrProGlyCysValProCysValArgGluAsnAsnSerSerArg 120
QY 361 TGCCTGGGTAGCGGTACCCCGACGCTGCGAGCTGGAACGCGACGCTCCCACTAGACA 420
Db 121 CysThrPValAlaLeuThrProThrLeuAlaIleArgAsnAlaSerValProThrThrThr 140
QY 421 ATACGACGCCACGTGATTTGCTGCTGTTGGGGGCGGCTCTTTCCTGCTATGTACGTG 480
Db 141 IleArgThrHisValAspLeuLeuValGlyAlaAlaIlePheCysSerAlaMetTyrVal 160
QY 481 GGGGATCTTGGCGGATGTGCTTCTCTGCTGCCAGCTGTTCACCAATCTGCGCCGG 540
Db 161 G1yAspLeuCysGlySerValPheLeuValSerGlnLeuPheThrIleSerProArgArg 180
QY 541 CATGAGAGCGGTGACGAGCTGCAATTTGCTCAATCATCCGGGCCAATAAGCGTCACGCT 600
Db 181 HisGluThrValGlnAspCysAsnCysSerIleTyrProGlyHisIleThrGlyHisArg 200
QY 601 ATGCTTGGGATATGATGATGAACCTGG 627
Db 201 MetAlaThrPAspMetMetMetAlaThrP 209

RESULT 15
US-10-020-510-6
; Sequence 6, Application US/10020510
; GENERAL INFORMATION:
; APPLICANT: Innogenetics N.V.
; TITLE OF INVENTION: Purified hepatitis C virus envelope proteins for diagnostic an
; TITLE OF INVENTION: therapeutic use.
; FILE REFERENCE: 2531-72
```

;; CURRENT APPLICATION NUMBER: US/10/020,510  
;; CURRENT FILING DATE: 2002-07-15  
;; NUMBER OF SEQ ID NOS: 122  
;; SOFTWARE: PatentIn 3.1  
;; SEQ ID NO 6  
;; LENGTH: 263  
;; TYPE: PRP  
;; ORGANISM: Hepatitis C virus  
US-10-020-510-6

## Alignment Scores:

Pred. No.:	1,79e-95	length:	263
Score:	1113.00	Matches:	207
Percent Similarity:	100.00%	Conservative:	2
Best Local Similarity:	99.04%	Mismatches:	0
Query Match:	96.36%	Indels:	0
DB:	24	Gaps:	0

US-09-899-303a-7 (1-633) x US-10-020-510-6 (1-263)

QY 1 ATGTTGGGTAAAGTCAATCGATACCTTAAGTGGGCTTCCGACCTCATGGGGTACATT 60  
|||  
Db 1 MetLeuGlyLysValIleAspThrLeuThrCysGlyPheAlaAspLeuValGlyTyrIle 20  
61 CCGCTGCTGGCGCCCGCCCTAGGGGGTCTGCCAGAGCCCTGGCGCATGGCGTCCGGTT 120  
|||  
Db 21 ProLeuValGlyAlaProLeuGlyGlyAlaAlaArgAlaLeuAlaHisGlyValArgVal 40  
QY 121 CTGGAAGACGGCGGTAACTATGCAACAGGAATTTGCTGTTGCTTTCTTCTATCTTC 180  
|||  
Db 41 LeuGluAspGlyValAsnTyrAlaThrGlyAsnLeuProGlyCysSerPheSerIlePhe 60  
QY 181 CTCTTGGCTTTACTGCTGCTGTGACCAATTCAGCTTCGCTTATGAGTGGCAACGTG 240  
|||  
Db 61 LeuLeuAlaLeuLeuSerCysLeuThrValProAlaSerAlaTyrGluValArgAsnVal 80  
QY 241 TCCGGGATGTACCATGTCAAGACGACTGCTCCAACTCAAGCATTTGTATGAGGACGG 300  
|||  
Db 81 SerGlyMetTyrHisValIleThrAsnAspCysSerAsnSerIleValIleTyrGluAlaIle 100  
QY 301 GACATGATCATGCACACCCCGGGGTGCGCTGCGCTCGGGAACAACCTTCCCGC 360  
|||  
Db 101 AspMetIleMetHisThrProGlyCysValProCysValArgGluAsnAsnSerSerArg 120  
QY 361 TGCTGGGTAGCGCTCAACCCCGACGCTCGAGCTAGGAACGCCAGCGTCCCGACTACGACA 420  
|||  
Db 121 CysTrpValAlaLeuThrProThrLeuAlaAlaArgAsnAlaSerValProThrThrThr 140  
QY 421 ATACGACGCCACGTCGATTTGCTGCTGGGGCGGCTGCTTTCTGTTCCGCTATGTACGTG 480  
|||  
Db 141 IleArgArgHisValAspLeuLeuValGlyAlaAlaAlaPheCysSerAlaMetTyrVal 160  
QY 481 GGGGATCTGTGGGATCTGCTCTCTCGTCTCCAGCTGTCACATCTCGCGCTCGCGG 540  
|||  
Db 161 GlyAspLeuCysGlySerValPheLeuValSerGlnLeuPheThrIleSerProAlaArg 180  
QY 541 CATGACAGCGGTGACGACTGCAATTGCTCAATCTATCCGGCCACATAACAGGTCAACGT 600  
|||  
Db 181 HisGluThrValGlnAspCysAsnSerIleTyrProGlyHisIleThrGlyHisArg 200  
QY 601 ATGGCTTGGGATGTGATGAAGAACTGG 627  
|||  
Db 201 MetAlaIlePaspMetMetMetAsnTrp 209

Search completed: February 12, 2003, 09:08:57  
Job time : 245 secs

GenCore version 5.1.3  
Copyright (c) 1993 - 2003 Compugen Ltd.

## OM nucleic - protein search, using frame\_plus\_n2p model

Run on: February 12, 2003, 08:55:00 ; Search time 61 Seconds

(without alignments)  
4276.322 Million cell updates/sec

Title: US-09-899-303a-7

Perfect score: 1155  
Sequence: 1 ATGTGGGTAGGTCATCGA.....TGATGATGAACGTGTAATAG 633

## Scoring table:

BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 1343160

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

## Command line parameters:

-MODEL=frame+ n2p.model -DEV=xiip  
-O=cpu2.1/uspro.apool/US09899303/runat\_12022003.085449.28268/app\_query.fasta.1.775  
-DB=SPREMBL.21 -QFMT=fastan -SUFFIX=n2p.rspt -MINMATCH=0.1 -LOOPCL=0  
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=biosum62 -TRANS=human40.cdi  
-LIST=45 -DOCALLIN=200 -THR.SCORE=pct -THR.MAX=100 -THR.MIN=0 -ALIGN=15  
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000  
-USER=US09899303 @CPU.1.1.88 @runat\_12022003.085449.28268 -NCPV=6 -ICPV=3  
-NO\_XLPHY -NO\_MMAP -LARGEQUERY -NEG\_SCORES=0 -WAIT -LONGLOG -DEV.TIMEOUT=120  
-MAIN.TIMEOUT=30 -THREDS=1 -XGAPOP=10 -XGAPEXT=0.5 -Fgapop=6 -Fgapext=7  
-Ygapop=10 -Ygapext=0.5 -Delop=6 -Delext=7

## Database :

SPREMBL.21:\*  
1: sp\_archaea:\*  
2: sp\_bacteria:\*  
3: sp\_fungi:\*  
4: sp\_human:\*  
5: sp\_invertebrate:\*  
6: sp\_mammal:\*  
7: sp\_mhc:\*  
8: sp\_organelle:\*  
9: sp\_phage:\*  
10: sp\_plant:\*  
11: sp\_podent:\*  
12: sp\_virus:\*  
13: sp\_vertebrate:\*  
14: sp\_unclassified:\*  
15: sp\_virus:\*  
16: sp\_bacteriaph:\*  
17: sp\_archaeap:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1095	94.8	310	Q90048	Q90048 hepatitis c

2	1094	94.7	513	12	081221	081221 hepatitis c
3	1094	94.7	3010	12	002829	002829 h genome po
4	1092	94.5	3010	12	09WXX2	09WXX2 hepatitis c
5	1091	94.5	3010	12	09J3H2	09J3H2 hepatitis c
6	1090	94.4	3010	12	09OP06	09OP06 hepatitis c
7	1089	94.3	3010	12	09J310	09J310 hepatitis c
8	1089	94.3	3010	12	09J367	09J367 hepatitis c
9	1086	94.0	3010	12	09OJY3	09OJY3 hepatitis c
10	1086	94.0	3010	12	081760	081760 hepatitis c
11	1085	93.9	3010	12	092972	092972 hepatitis c
12	1085	93.9	3010	12	09OJY4	09OJY4 hepatitis c
13	1084	93.9	441	12	081776	081776 hepatitis c
14	1084	93.9	802	12	068520	068520 hepatitis c
15	1084	93.9	3010	12	09J3H7	09J3H7 hepatitis c
16	1084	93.9	3010	12	09J3H7	09J3H7 hepatitis c
17	1083	93.8	587	12	068518	068518 hepatitis c
18	1083	93.8	787	12	008244	008244 h genome po
19	1083	93.8	1008	12	089156	089156 hepatitis c
20	1083	93.8	3010	12	09D7D6	09D7D6 hepatitis c
21	1083	93.8	3013	12	09OJY0	09OJY0 hepatitis c
22	1083	93.8	3010	12	09OJY9	09OJY9 hepatitis c
23	1082	93.7	3010	12	09D7D9	09D7D9 hepatitis c
24	1082	93.7	3010	12	09J3H9	09J3H9 hepatitis c
25	1082	93.7	3010	12	09J3H5	09J3H5 hepatitis c
26	1082	93.7	3010	12	08OJL8	08OJL8 hepatitis c
27	1081	93.6	337	12	086880	086880 hepatitis c
28	1081	93.6	2864	12	09WJL8	09WJL8 hepatitis c
29	1081	93.6	2864	12	092974	092974 hepatitis c
30	1081	93.6	3010	12	08V638	08V638 hepatitis c
31	1081	93.6	3010	12	09D7E4	09D7E4 hepatitis c
32	1081	93.6	3010	12	092969	092969 hepatitis c
33	1081	93.6	3010	12	09OJY8	09OJY8 hepatitis c
34	1081	93.6	3010	12	09OJY7	09OJY7 hepatitis c
35	1081	93.6	3010	12	09OJY8	09OJY8 hepatitis c
36	1081	93.6	3010	12	09OJY7	09OJY7 hepatitis c
37	1081	93.6	3010	12	09J1A0	09J1A0 hepatitis c
38	1081	93.6	3010	12	09J1V3	09J1V3 hepatitis c
39	1080	93.5	2864	12	09WJL9	09WJL9 hepatitis c
40	1080	93.5	3010	12	092971	092971 hepatitis c
41	1080	93.5	3010	12	002828	002828 h genome po
42	1080	93.5	3012	12	09WJL7	09WJL7 hepatitis c
43	1079	93.4	1008	12	089157	089157 hepatitis c
44	1079	93.4	2864	12	092975	092975 hepatitis c
45	1079	93.4	3008	12	09J3F4	09J3F4 hepatitis c

## ALIGNMENTS

RESULT 1  
Q90048 PRELIMINARY: PRT: 310 AA.  
AC Q90048;  
DT 01-NOV-1996 (TREMUREL.01, Last sequence update)  
DT 01-NOV-1996 (TREMUREL.20, Last annotation update)  
DE Genome polypeptide [Contains: envelope glycoprotein E2 (GP68) (GP70)  
(NS1)] (Fragment).  
GN POLYPEPTIDE.  
OS Hepatitis C virus.  
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;  
OC Hepacivirus.  
OX NCBI\_TaxID:11103;  
RN [1]  
RP MEDLINE:93019054; PubMed:1328487;  
RA Kohara M., Tsukiyama-Kohara K., Maki N., Asano K., Yamaguchi K.,  
RA Miki K., Tanaka S., Hattori N., Matsunaga Y., Saito I., et al;  
RT "Expression and characterization of glycoprotein gp35 of hepatitis C  
virus using recombinant vaccinia virus.";  
RL J. Gen. Virol. 73:2313-2318(1992).  
CC -!- SIMILARITY: TO HEPATITIS C VIRUS ENVELOPE GLYCOPROTEIN E1.  
DR EMBL: S46012; AAB23552.1; -;  
InterPro: IPR002521; HCV\_core.

162 (67)

ID	Accession	Protein Name	Length (aa)	Score	Percent Similarity	Best Local Similarity	Query Match	DB
DR	InterPro: IPR002519; HCV Env.							
DR	InterPro: IPR002531; HCV NS1.							
DR	Pfam: PF01542; HCV_core; 1.							
DR	Pfam: PF01538; HCV_env; 1.							
DR	Pfam: PF01560; HCV_NS1; 1.							
DR	Prodom: PD186062; HCV_NS1; 1.							
KM	Coat protein; Envelope protein; Glycoprotein; Nonstructural protein; Polyprotein; Transmembrane.							
FT	NON_TER							
SO	SEQUENCE	310 AA; 33200 MW; 5DB04CE7E70E41F9 CRC64;						
Alignment Scores:								
QY	4	TTGGGTAAGTCATCGATGCAATACCCCTTACGTCGGCTTCGCCGACCTCATGGGTAACATCCG	310	2.35e-97	1095.00	98.56%	97.60%	94.81%
Db	3	LeuLIeLysValLIeLAspThrLeuThrCysGILpHeLaAspLeuMetGILyTyrLePro	203					
QY	64	CTCGCGGGCGCCCGCCATAGGGGGGTCGCGAAGCCCTGGCGCATGGCGCCGGCTTCG	123					
Db	23	LeuValGILyLaProLeuGILyGILyLaLaLaLaLaLaLaLaLaLaLaLaLaLaLaLaLaLaLa	42					
QY	124	GAAGACGGCGGTAACATGCAACAGGGAATTTGGCTGGTGGCTCTTCTATCTTCCTC	183					
Db	43	GILuAspGILyLa	62					
QY	184	TTGGCTTACTGTCGCTGTCTGACCATTCACAGCTTCGCTTATGAGGTGCGCAAGTGTCC	243					
Db	63	LeuLaLeuLeuSerCysLeuThrILeProLaSerLaTyrGILyLaLaLaLaLaLaLaLaSer	82					
QY	244	GGGATGTAACATGTCACAGAGAGATGTCGCAACATTCGTGATAGCGACGGGAC	303					
Db	83	GILyLeuThrLysValLThrAsnAspCysSerLAsnSerLLeuValTyrGILyLaLaLaLaLa	102					
QY	304	ATGATCATGACACACCCCGGGGCGGTGCGCTGCGTGGGAGAAACAATCTTCCCGCTGC	363					
Db	103	MetLeuMetLThrProGILyCysValProCysValLaLaGILyLaLaLaLaLaLaLaLaLa	122					
QY	364	TTGGTAGCGCTCACCCCGACGCTGCGACGTGAGAACCCAGCGTCCCGCATACGACATA	423					
Db	123	ThrLaLaLeuThrProThrLeuLaLaLaLaLaLaLaLaLaLaLaLaLaLaLaLaLaLaLa	142					
QY	424	CGAGGCAACGTCGATTTGTCGTTGGGGGGCGGCTTTCGTTCGGTAACTAGTACGTGGG	483					
Db	143	ArgThrGILyLaLaLaLeuLeuValGILyLaLaLaLaLaLaLaLaLaLaLaLaLaLaLaLa	162					
QY	484	GATTCGCGGATGTCCTTCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG	543					
Db	163	AspLeuYsgILySerValRheLeuValSerGILeLeuThrRheSerProLaGILyGILy	182					
QY	544	GAGACGGTGCAGACATTCGTCATGTATCCGCGACATTAACAGTCAACGCTATG	603					
Db	183	GILuThrValGILeAsnAspCysAsnAspSerLLeuYsgILyLeuThrGILyLaLaLaMet	202					
QY	604	GCTTGGGATATGATGATGACTGG	627					
Db	203	AlaThrAspMetMetMetLAsnTyr	210					
RESULT 2								
ID	081221	PRELIMINARY:	PRT:	513	AA.			
AC	081221;							
DT	01-NOV-1996 (TRMBLrel. 01, Created)							
DT	01-NOV-1996 (TRMBLrel. 01, Last sequence update)							
DT	01-DEC-2001 (TRMBLrel. 19, Last annotation update)							
DE	Genome polyprotein [Contains: envelope glycoprotein E2 (GP68) (GP70)							

```

DE (NS1)) (Fragment).
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID:11103;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-HC-J4;
RX MEDLINE=91013116; PubMed=2170712;
RA Okamoto H., Okada S., Sugiyama Y., Yotsumoto S., Tanaka T.,
RA Yoshizawa H.;
RT "The 5'-terminal sequence of the hepatitis C virus genome.";
RL Jpn. J. Exp. Med. 60:167-177(1990).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-HC-J4;
RX MEDLINE=92044440; PubMed=1658196;
RA Okamoto H., Okada S., Sugiyama Y., Kural K., Iizuka H., Machida A.,
RA Miyakawa H., Mayumi M.;
RT "Nucleotide sequences of the genomic RNA of hepatitis C virus isolated
RT from a human carrier: comparison with reported isolates for conserved
RT and divergent regions.";
RL J. Gen. Virol. 72:2697-2704(1991).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN-HC-J4;
RX MEDLINE=94270990; PubMed=7545932;
RA Hotta H., Doi H., Hayashi T., Purwanta M., Soemarto W., Mizokami M.,
RA Ohba K., Homma M.;
RT "Analysis of the core and E1 envelope region sequences of a novel
RT variant of hepatitis C virus obtained in Indonesia.";
RL Arch. Virol. 136:53-62(1994).
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN-HC-J4;
RX MEDLINE=94174722; PubMed=7510436;
RA Mink M., Benichou S., Madaule P., Tiollais P., Prince A.,
RA Inchausti G.;
RT "Characterization and mapping of a B-cell immunogenic domain in
RT hepatitis C virus E2 glycoprotein using a yeast peptide library.";
RL Virology 200:246-255(1994).
CC -1- SIMILARITY: TO: HEPATITIS C VIRUS ENVELOPE GLYCOPROTEIN E1.
DR EMBL: D00833; BAA00706.1; -
DR InterPro: IPR002522; HCV_capsid.
DR InterPro: IPR002521; HCV_core.
DR InterPro: IPR002519; HCV_env.
DR InterPro: IPR002531; HCV_NSI.
DR Pfam: PF01543; HCV_capsid.1.
DR Pfam: PF01542; HCV_core.1.
DR Pfam: PF01539; HCV_env.1.
DR Pfam: PF01560; HCV_NSI.1.
DR ProDom: PD186062; HCV_NSI.1.
KW Coat protein; Envelope protein; Glycoprotein; Nonstructural protein;
KW Polypeptid; Transmembrane.
RN NON_TER 513
SO SEQUENCE 513 AA; 35874 MW; 3C8E1D94FD1FA7E2 CRC64;

Alignment Scores:
Pred. NO.: 3.15e-97 Length: 513
Score: 1094.00 Matches: 203
Percent Similarity: 99.52% Conservative: 4
Best Local Similarity: 97.60% Mismatches: 1
Query Match: 94.72% Indels: 0
DB: 12 Gaps: 0

US-09-899-303A-7 (1-633) x Q81221 (1-513)
OY 4 TTGGGTAAAGTCATCGATACCTTACGTGCGGCTTCCGACGACCTCATGGGTACATTCG 63
Db 119 LcnglyysvallileasprhrlautnrcysglyPhelialaspneuwetcltyrrilepro 138
OY 64 CTCGTGGCGGCCCCCTTAGGGGGGTGCTGCGCAGAGCCCTTGCGGCGATGCGTCCGGGTTCTG 123

```

[illegible]

Accession	Protein	Length	Score	Percent Similarity	Best Local Similarity	Query Match	DB
US-09-899-303A-7 (1-633) x Q02829 (1-3010)							
QY	4	TTGGGTAAGGCAATGATGATACCGCTTACGGGGCGCTTGGCCAGCATGATGGGTACATTCGG	63				
Db	119	LeuValGlyValIleAspPThrLeuThrCysGlyPheAlaAspLeuMetGlyTyrIlePro	138				
QY	64	CTCTCGGCGCCGCCCTAAGGGGGTGTCTGCAGACCCCTGGCGCATGCGTCCGGGTTCTG	123				
Db	139	LeuValGlyAlaProLeuGlyGlyAlaAlaArgAlaLeuAlaHisGlyValAlaGlyValLeu	158				
QY	124	GAGAGCGCGGTGAACATATGCAACAGGGAATTTGGCTGTCTTCTCTATCTTCTC	183				
Db	159	GluAspGlyValAlaSerTyrAlaThrGlyAsnLeuProGlyCysSerPheSerIlePheLeu	178				

OY	184	TTCGGTTACATGCTCGTGTCTGACCATTCACACTTCGCCTTAATAGAGGCCAAGCTGC	243
Db	179	LeuAlaIeuleuSerCysIeuThiIleProAlaSerIatIyGIuValaIagAsnValSer	1988
OY	244	GGAATGTACCAATGTCACAGAACAGACTGCTCCAACTCAAGACAATTGTATGAGGCAAGCGAC	3030
Db	199	GLYTILETHINISVALTHRASPcysSerSernSerIleValTYIGLUALaIaIaAsp	218
OY	304	ATGATTCATGACACCACCCCGGGTGCGTCCCTGGCTTGGGGAGAACAACACTCTCCCGCTGC	3635
Db	219	MettIeMetHisThrProGLYCysValaIProCysValaIArgGIuaSpasnsSerSetaArgCys	238
OY	364	TGGGATGAGCGTCACCCCCCACCGCTGCGAGCTGAGAACCCACAGCGCCCACTAGACAATA	4233
Db	239	TrpValaIaIeuthrProthIreuAlaIaIaArgsnlaIaSerValaIprothIrThrThIle	258
OY	424	CGAGCCACAGTCGATTTGCTCGTTGGGGCGGCGCTGCTGTCTGCTCCGATATGACGTGGGG	4833
Db	259	ArgaIrgHisValaIaSplueIeUvalaIGLYaIaIaIaIaIheCysSerIaIameTyuValaIGly	278
OY	484	GATCTCTCGGATCTGTCTCTCTGCTCCAGCTGTTCACACATCTGCGCTCGCGGCAAT	5433
Db	279	AspleucysGIyseValaIpheIeValaSerIInleuethIrPheSerProaIdaIghis	298
OY	544	GAGACGGTGCAGCACTGCAATTCCTCAATCTATCCCGGACACATAAGGTCACCGTATG	6030
Db	299	GIuThrValaIGlnaIaRspCysAsnCysSerIleIyrProGIuHIsIeUsenGIuHIsArgMet	318
OY	604	GCTTGGGATATGATGATGACACTGG	627
Db	319	Alatrpaspmetmetmetasntrp	326
<hr/>			
RESULT 4			
O9WMX2		PRELIMINARY:	PRT; 3010 AA.
ID	O9WMX2		
AC	O9WMX2;		
DT	01-NOV-1999 (TREMBLrel. 12, Created)		
DD	01-NOV-1999 (TREMBLrel. 12, Last sequence update)		
DT	01-JUN-2002 (TREMBLrel. 21, Last annotation update)		
DE	Genome polypeptide [Contains: envelope glycoprotein E2 (GP68) (GP70) (NS1)].		
DE	Hepatitis C virus type 1b.		
OS	Viruslike ssRNA positive-strand viruses, no DNA stage; Flaviviridae;		
OC	Hepadnavirus.		
OX	NCBI_TaxID=31647;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN-RB:		
RX	Lojmann V., Koerner F., Koch J.O., Herian U., Theilmann L.,		
RA	Bartenschlager R.;		
RT	"Replication of subgenomic hepatitis C virus RNAs in a hepatoma cell		
RL	line";		
Science 285:110-113(1999).			
CC	-1- SIMILARITY: TO HEPATITIS C VIRUS ENVELOPE GLYCOPROTEIN EI.		
DR	EMBL; AJ238799; CAB46677.1; -;		
DR	HSSP; P26663; IXP.		
DR	MEROPS; S29_001; -;		
DR	InterPro; IPR001410; DEAD.		
DR	InterPro; IPR002522; HCV_capsid.		
DR	InterPro; IPR002521; HCV_core.		
DR	InterPro; IPR002519; HCV_env.		
DR	InterPro; IPR002531; HCV_NS1.		
DR	InterPro; IPR002518; HCV_NS2.		
DR	InterPro; IPR004109; HCV_NS3.		
DR	InterPro; IPR000745; HCV_NS4a.		
DR	InterPro; IPR001490; HCV_NS4b.		

DR	InterPro: IPR002868; HCV_NS5.
DR	InterPro: IPR002166; HCV_RDRP.
DR	Pfam: PF01543; HCV_capsid. 1.
DR	Pfam: PF01542; HCV_core. 1.
DR	Pfam: PF01539; HCV_env. 1.
DR	Pfam: PF01560; HCV_NS1. 1.
DR	Pfam: PF01538; HCV_NS2. 1.
DR	Pfam: PF02907; HCV_NS3. 1.
DR	Pfam: PF01006; HCV_NS4a. 1.
DR	Pfam: PF01001; HCV_NS4b. 1.
DR	Pfam: PF01506; HCV_NS5a. 1.
DR	Pfam: PF00998; HCV_RDRP. 1.
DR	Prodom: PD186062; HCV_NS1. 1.
KW	Anti-binding; Coat protein; Envelope protein; Glycoprotein; Helicase; Nonstructural protein; Polypotein; RNA-directed RNA polymerase;
KW	Transmembrane.
KM	CHAIN 1 191 CORE PROTEIN.
FT	CHAIN 192 383 GLYCOPROTEIN E1.
FT	CHAIN 384 746 GLYCOPROTEIN E2.
FT	CHAIN 747 809 P7 PROTEIN.
FT	CHAIN 810 1026 NS2 PROTEINASE.
FT	CHAIN 1027 1657 NS3 PROTEINASE/HELICASE.
FT	CHAIN 1658 1711 NS3/4A PROTEINASE COFACTOR.
FT	CHAIN 1712 1972 NS4B PROTEIN.
FT	CHAIN 1973 2419 NS5A PHOSPHOPROTEIN.
FT	CHAIN 2420 3010 NS5B RNA DEPENDANT RNA POLYMERASE.
QO	SEQUENCE 3010 AA: 326902 MW: 37B3448DAFA9A10B CRC64:

**Alignment Scores:**

Pred. No.:	6.2/6-97	Length:	3010
Score:	1092.00	Matches:	202
Percent Similarity:	99.04%	Conservative:	4
Best Local Similarity:	97.12%	Mismatches:	2
Query Match:	94.55%	Indels:	0
B:	12	Gaps:	0

US-09-899-303A-7 (1-633) x Q9WMX2 (1-3010)

QY	4	TTGGGTAAAGGCATCGATTACCCCTACCTAGTGGCGGCTTCCGACCTCATAGGGGTACATTCGC	63
Db	119	LeuLIyysValIIeAspThrLeuThrCysclYphelAlaspIeuMetGIYIlePro	138
QY	64	CTCTGGGCGCCCCCTTAGGGGGGTGCTGCCAGAGCCCTGGCGCATGCGCTCCGGGTTCTG	123
Db	139	LeuValIGlyAlaProIeuGIgYlAlaAlaArgAlaIleuAlaHISGlyValArgValIleu	158
QY	124	GAATACGCGCGTAATATGCACAGGGGAATTGGCTGGTGCCTTTCTCTATCTTCCTC	189
Db	159	GIuAspGIyAlaAsnTYrAlaThrGIyAlaAsnIeuProLIyCysSerPheSerIlePheIeu	178
QY	184	TTGGCTTTACTGTCTGTCTGACCATTCACAGCTTCGCGCTTATGAGGTGGCGCAAGTGTCC	243
Db	179	LeuAlaIeuIeuSerCysLeuThrIleProAlaSerIatYrGIuValArgAsnValSer	198
QY	244	GGATGTACCACTGTACAGAACAGATGTGTCACATCGACAGANTTGCTGTAGAGCGACGGAC	303
Db	199	GIyAlaIYrHISValIIrHnAsnAspCysSerAlaSerIleValYrGIuAlaAlaAsp	218
QY	304	ATGATCATGCACACCCCCGGGTGGTCCCTGTCTGGGAGAACATCTTCCCGCTGC	363
Db	219	MetIleMetHISThrProGIyCysValProCysValArgGIuAsnAsnSerSerArgCys	238
QY	364	TGGGTACCGCTCACCCCCAGGCTTGCAAGCTTAGAAGCGCAGCGTCCCACTAGACACATA	422
Db	239	TrpValAlaIeuThrProThrIleuAlaAlaArgAsnAlaSerValProThrThrIle	258
QY	424	CGAGGCACACGCGATTGTGCTGTGGGGGGCGGCTTCTTGTTCCGCTATGTACGTGGG	483
Db	259	ArgArgHISValAspIeuIeuValGIyAlaAlaIalaIalaIeuCysSerAlaIleMetYrValGIy	278
QY	484	GATCTTGCGGATCTGTCTTCCTGCTGCCAGCGTGTACCAACATCTGCGCTCGCGGCGAT	543
Db	279	AspIeuCysGIySerValPheIeuValAlGIuIleuPheThrPheSerProArgHIS	298



DR InterPro: IPR001410; DEAD.  
 DR InterPro: IPR002522; HCV\_capsid.  
 DR InterPro: IPR002521; HCV\_core.  
 DR InterPro: IPR002519; HCV\_env.  
 DR InterPro: IPR002531; HCV\_NS1.  
 DR InterPro: IPR002518; HCV\_NS2.  
 DR InterPro: IPR004109; HCV\_NS3.  
 DR InterPro: IPR00745; HCV\_NS4.  
 DR InterPro: IPR001490; HCV\_NS4b.  
 DR InterPro: IPR002868; HCV\_NS5a.  
 DR InterPro: IPR002166; HCV\_RdRP.  
 DR Pfam: PF01543; HCV\_capsid; 1.  
 DR Pfam: PF01542; HCV\_core; 1.  
 DR Pfam: PF01539; HCV\_env; 1.  
 DR Pfam: PF01560; HCV\_NS1; 1.  
 DR Pfam: PF01538; HCV\_NS2; 1.  
 DR Pfam: PF02907; HCV\_NS3; 1.  
 DR Pfam: PF01006; HCV\_NS4a; 1.  
 DR Pfam: PF01001; HCV\_NS4b; 1.  
 DR Pfam: PF01506; HCV\_NS5a; 1.  
 DR Pfam: PF00998; HCV\_RdRP; 1.  
 DR ProDom: PD186062; HCV\_NS1; 1.  
 DR ATP-binding: Coat protein; Envelope protein; Glycoprotein; Helicase;  
 KM Nonstructural protein; Polyprotein; RNA-directed RNA polymerase;  
 KM Transmembrane.  
 FT CHAIN 1 191 CORE PROTEIN,  
 FT CHAIN 192 383 GLYCOPROTEIN E1.  
 FT CHAIN 384 746 GLYCOPROTEIN E2.  
 FT CHAIN 747 809 P1 PEPTIDE.  
 FT CHAIN 810 1026 NS2 PROTEINASE.  
 FT CHAIN 1027 1657 NS3 PROTEINASE/HELICASE.  
 FT CHAIN 1658 1711 NS4/4A PROTEINASE COFACTOR.  
 FT CHAIN 1712 1972 NS4B PROTEIN.  
 FT CHAIN 1973 2418 NS5A PHOSPHOPROTEIN.  
 FT CHAIN 2419 3010 NS5B RNA DEPENDANT RNA POLYMERASE.  
 SQ SEQUENCE 3010 AA; 326999 MW; A570BB980DD64634 CRC64;

## Alignment Scores:

Pred. No.: 9.79e-97 Length: 3010  
 Score: 1090.00 Matches: 201  
 Percent Similarity: 99.04% Conservative: 5  
 Best Local Similarity: 96.63% Mismatches: 2  
 Query Match: 94.37% Indels: 0  
 DB: 12 Gaps: 0

US-09-899-303A-7 (1-633) x Q9P06 (1-3010)

QY 4 TTGGGTAAAGTATGATGATACCCCTTACGTGCGGCTTCCGACCTCATGGGTTACATTCG 63  
 DB 119 LeuGlyLysValIleAspThrLeuThrCysGlyPheAlaAspLeuMetGlyTyrLeuPro 138  
 QY 64 CTCGTGCGGCCCCCTAGGGGGTGTGCCAGAGCCCTGGCGATGGCTCCGGTTCG 123  
 DB 139 LeuValIGlyAlaProLeuGlyGlyAlaAlaArgAlaLeuAlaHisGlyValAlaGlyValLeu 158  
 QY 124 GAAGACGGCGGTAACTATGCAACAGGGAATTCCTGGTTCCTTCTATCTTCCTC 183  
 DB 159 GluAspGlyValAlaAsnTyrIleThrGlyAlaLeuProGlyCysSerHisSerIlePheLeu 178  
 QY 184 TTGGCTTACCTGCTGCTGTGACCAATTCAGCTCCGTTAGAGGTGGCAACGTCCTC 243  
 DB 179 LeuAlaLeuLeuSerCysLeuThrIleProAlaSerAlaTyrGlyValAlaIleAsp 198  
 QY 244 GGGATGACCATGTCACGAGACGATGCTCAACTCAAGCATTTGTATGAGGCGGAGC 303  
 DB 199 GlyValTyrHisValThrAsnAspCysSerAlaAlaSerIleValTyrGlyValAlaIleAsp 218  
 QY 304 ATGATCATGACACACCCCGGGTGGTCCCTGGCTTCGGGAGAACACATCTTCCCGTCG 363  
 DB 219 MetIleMetHisIleThrProGlyCysValProCysValAlaGlyAlaAsnAsnSerSerArgCys 238  
 QY 364 TGGGTAGCGCTACCCCGACGCGGCTGAGTAGAAGCGCAGCTGCCCACTACGACAATA 423

DB 239 TyrValAlaLeuThrProThrIleuAlaAlaArgAsnAlaSerValProThrThrIle 258  
 QY 424 CGACGCCACGTGATGTCGTCGTTGGGGCGGTCTTCTGTTCCGCTATGATGAGGGG 483  
 DB 259 ArgArgHisValAspLeuLeuValGlyAlaAlaAlaLeuCysSerAlaMetTyrValGly 278  
 QY 484 GATCTCTGCGGATCTGCTCTTCTGTCCTCCAGCTGTTACCATCTCGGCTCGCGGCAT 543  
 DB 279 AspLeuCysGlySerValPheLeuValAlaGlnLeuPheThrPheSerProArgArgHis 298  
 QY 544 GAGACGGTGCAGACGCAATGCTCAATCTATCCGGCGCACATACAGCTACCGCTATG 603  
 DB 299 GlyThrValGlnAspCysAsnGlySerIleTyrProGlyHisValThrGlyHisArgMet 318  
 QY 604 GCTTGGATATGATGATGATGATG 627  
 DB 319 AlaTrpAspMetMetMetAsnTrp 326  
 RESULT 7  
 Q9J310 PRELIMINARY; PRT; 3010 AA.  
 AC Q9J310;  
 DT 01-OCT-2000 (TREMBLrel. 15, Created)  
 DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)  
 DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)  
 DE Genome polyprotein [Contains: envelope glycoprotein E2 (GP68) (GP70) (NS1)].  
 DE Hepatitis C virus.  
 OS Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;  
 OC Hepacivirus.  
 NCBI\_Taxid=11103;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=MD12;  
 RA Nagayama K., Kurosaki M., Enomoto N., Miyasaka Y., Marumo F., Sato C.;  
 RT "Characteristics of hepatitis C viral genome associated with disease progression."  
 RL Submitted (NOV-1999) to the EMBL/Genbank/DBJ databases.  
 CC -1- SIMILARITY: TO HEPATITIS C VIRUS ENVELOPE GLYCOPROTEIN E1.  
 DR EMBL: AF207753; AAF65943.1; -.  
 DR HSSP: P26663; 1JXP.  
 DR MEROPS: S29.001; -.  
 DR MEROPS: U39.001; -.  
 DR InterPro: IPR000345; Cytochrome\_bind.  
 DR InterPro: IPR001410; DEAD.  
 DR InterPro: IPR002522; HCV\_capsid.  
 DR InterPro: IPR002521; HCV\_core.  
 DR InterPro: IPR002519; HCV\_env.  
 DR InterPro: IPR002531; HCV\_NS1.  
 DR InterPro: IPR002518; HCV\_NS2.  
 DR InterPro: IPR004109; HCV\_NS3.  
 DR InterPro: IPR00745; HCV\_NS4.  
 DR InterPro: IPR001490; HCV\_NS4b.  
 DR InterPro: IPR002868; HCV\_NS5a.  
 DR InterPro: IPR002166; HCV\_RdRP.  
 DR InterPro: IPR001650; Helicase\_C.  
 DR Pfam: PF01543; HCV\_capsid; 1.  
 DR Pfam: PF01542; HCV\_core; 1.  
 DR Pfam: PF01539; HCV\_env; 1.  
 DR Pfam: PF01560; HCV\_NS1; 1.  
 DR Pfam: PF01538; HCV\_NS2; 1.  
 DR Pfam: PF02907; HCV\_NS3; 1.  
 DR Pfam: PF01006; HCV\_NS4a; 1.  
 DR Pfam: PF01001; HCV\_NS4b; 1.  
 DR Pfam: PF01506; HCV\_NS5a; 1.  
 DR Pfam: PF00998; HCV\_RdRP; 1.  
 DR Pfam: PF00271; Helicase\_C; 1.  
 DR ProDom: PD186062; HCV\_NS1; 1.  
 DR PROSITE: PS00190; CYTOCHROME\_C; UNKNOWN\_1.  
 DR Coat protein; Envelope protein; Glycoprotein; Nonstructural protein;  
 KM Polyprotein; RNA-directed RNA polymerase; Transmembrane.  
 SQ SEQUENCE 3010 AA; 326692 MW; 074098DB305AFA1A9 CRC64;

## Alignment Scores:

Pred. No.: 1,22e-96 Length: 3010  
 Score: 1089.00 Matches: 201  
 Percent Similarity: 98.08% Conservative: 3  
 Best Local Similarity: 96.63% Mismatches: 4  
 Query Match: 94.29% Indels: 0  
 DB: 12 Gaps: 0

US-09-899-303a-7 (1-633) x Q9J3G7 (1-3010)

```

QY 4 TTGGGTAAAGTCAATGATACCTTACGTGGCGGCTTGGCCGACCTCATGGGGTACATTCCG 63
    |||||||
Db 119 LeuGlyLysValIleAspThrLeuThrCysGlyPheAlaAspLeuMetGlyTyrIlePro 138
QY 64 CTCGTGGCGCCCCCTAGAGGGGTGTCACAGACCCTGGCGCATGGCGGTCTGTG 123
    |||||||
Db 139 LeuValGlyAlaProLeuGlyGlyAlaAlaArgValLeuAlaHisGlyValArgValLeu 158
QY 124 GAAGACGGCGGTGAACATATGCAACAGGCAATTTGCTGCTGTTCTCTATCTTCCTC 183
    |||||||
Db 159 GluAspGlyValAsnTyrAlaThrGlyAsnLeuProGlyCysSerPheSerIlePheLeu 178
QY 184 TTGGCTTACTGCTGCTGTGACCATTCAGCTTCGCTATGAGGTCGACAGCTGCTCC 243
    |||||||
Db 179 LeuAlaLeuLeuSerCysLeuThrThrProAlaSerAlaTyrGlyValArgAsnValSer 198
QY 244 GGGATGATACCATGTCACAGAACGACTGCTCCAACTCAAGCATTTGTATGAGGACGCGAC 303
    |||||||
Db 199 GlyMetCysTrpHisValThrAsnAspCysSerAsnSerSerIleValIlyrGlyAlaAlaAsp 218
QY 304 ATGATCATGCAACACCCCGGGGTGCTGCTGCTGCTGCGGAGAACAACTTTCCCGCTGC 363
    |||||||
Db 219 MetIleMetHisIleAspProGlyCysValProCysValArgGlyAsnAsnSerSerArgCys 238
QY 364 TTGGTAGCGCTGACCCCGACGCTGCGAGCTAGGAAGCGACGCGCTCCACTAGGACATA 423
    |||||||
Db 239 TrpValAlaLeuThrProThrLeuAlaAlaArgAsnAlaSerValProThrThrIle 258
QY 424 CGAGCGCACGTCGATTTGCTGCTTGGGGCGGCTGCTTTCTGTTCCGCTATGACGTGGG 483
    |||||||
Db 259 ArgArgHisValAspLeuLeuValGlyAlaAlaAlaPheCysSerIleAlaMetIlyrValGly 278
QY 484 GATCTGCGCGGATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 543
    |||||||
Db 279 AspLeuCysGlySerValPheLeuIleSerGlyLeuPheThrPheSerProArgArgHis 298
QY 544 GAGACGGTGAAGACATGCTCAATCTCAATCCGCGGACCAATACAGTCAAGCTGATG 603
    |||||||
Db 299 GluThrValGlnAspCysAsnCysSerIleTyrProGlyHisValSerGlyHisArgMet 318
QY 604 GCTTGGGATATGATGATGAAGTGG 627
    |||||||
Db 319 AlaTrpAspMetMetMetAsnTrp 326

RESULT 8
Q9J3G7 PRELIMINARY: PRT: 3010 AA.
AC Q9J3G7:
DT 01-OCT-2000 (TREMBLrel. 15, Created)
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE Genome polyprotein [Contains: envelope glycoprotein E2 (GP68) (GP70)
DE (NS1)].
OS Hepatitis C virus.
OC Viruses: ssRNA positive-strand viruses, no DNA stage: Flaviviridae;
OC Hepcivirus
NCBI_TaxID=11103:
RX NCBI_TaxID=11103:
RN (1)
RP SEQUENCE FROM N.A.
RC STRAIN=MD25;
RA Nagayama K., Kurosaki M., Enomoto N., Miyasaka Y., Marumo F., Sato C.;
RT "Characteristics of hepatitis C viral genome associated with disease
RT progression.";
  
```

RL Submitted (NOV-1999) to the EMBL/GenBank/DBJ databases.  
 CC -1- SIMILARITY: TO HEPATITIS C VIRUS ENVELOPE GLYCOPROTEIN E1.  
 DR EMBL: AF207766; AAF65956.1; -  
 DR HSP: P26663; LUXP.  
 DR MEROPS: S29.001; -  
 DR MEROPS: U39.001; -  
 DR InterPro: IPR000345; Cytic\_heme\_bind.  
 DR InterPro: IPR001410; DEAD.  
 DR InterPro: IPR002522; HCV\_capsid.  
 DR InterPro: IPR002521; HCV\_core.  
 DR InterPro: IPR002519; HCV\_env.  
 DR InterPro: IPR002531; HCV\_NS1.  
 DR InterPro: IPR002518; HCV\_NS2.  
 DR InterPro: IPR004109; HCV\_NS3.  
 DR InterPro: IPR000745; HCV\_NS4a.  
 DR InterPro: IPR001490; HCV\_NS4b.  
 DR InterPro: IPR002868; HCV\_NS5a.  
 DR InterPro: IPR002166; HCV\_RdRP.  
 DR InterPro: IPR001650; Helicase\_C.  
 DR Pfam: PF01543; HCV\_capsid. 1.  
 DR Pfam: PF01539; HCV\_env. 1.  
 DR Pfam: PF01560; HCV\_NS1. 1.  
 DR Pfam: PF01538; HCV\_NS2. 1.  
 DR Pfam: PF02907; HCV\_NS3. 1.  
 DR Pfam: PF01006; HCV\_NS4a. 1.  
 DR Pfam: PF01506; HCV\_NS4b. 1.  
 DR Pfam: PF00998; HCV\_NS5a. 1.  
 DR Pfam: PF00271; Helicase\_C. 1.  
 DR ProDom: PD186062; HCV\_NS1. 1.  
 DR PROSITE: PS00190; CYTOCHROME\_C. UNKNOWN\_1.  
 KW Coat protein: Envelope protein; Glycoprotein; Nonstructural protein;  
 KW polyprotein; RNA-directed RNA polymerase; Transmembrane.  
 SQ SEQUENCE 3010 AA: 326507 MW: 98EAC8ADA4B44IDF CRC64:

## Alignment Scores:

Pred. No.: 1,22e-96 Length: 3010  
 Score: 1089.00 Matches: 202  
 Percent Similarity: 98.56% Conservative: 3  
 Best Local Similarity: 97.12% Mismatches: 3  
 Query Match: 94.29% Indels: 0  
 DB: 12 Gaps: 0

US-09-899-303a-7 (1-633) x Q9J3G7 (1-3010)

```

QY 4 TTGGGTAAAGTCAATGATACCTTACGTGGCGGCTTGGCCGACCTCATGGGGTACATTCCG 63
    |||||||
Db 119 LeuGlyLysValIleAspThrLeuThrCysGlyPheAlaAspLeuMetGlyTyrIlePro 138
QY 64 CTCGTGGCGCCCCCTAGAGGGGTGTCACAGACCCTGGCGCATGGCGGTCTGTG 123
    |||||||
Db 139 LeuValGlyAlaProLeuGlyGlyAlaAlaArgValLeuAlaHisGlyValArgValLeu 158
QY 124 GAAGACGGCGGTGAACATATGCAACAGGCAATTTGCTGCTGTTCTCTATCTTCCTC 183
    |||||||
Db 159 GluAspGlyValAsnTyrAlaThrGlyAsnLeuProGlyCysSerPheSerIlePheLeu 178
QY 184 TTGGCTTACTGCTGCTGTGACCATTCAGCTTCGCTATGAGGTCGACAGCTGCTCC 243
    |||||||
Db 179 LeuAlaLeuLeuSerCysLeuThrThrProAlaSerAlaTyrGlyValArgAsnValSer 198
QY 244 GGGATGATACCATGTCACAGAACGACTGCTCCAACTCAAGCATTTGTATGAGGACGCGAC 303
    |||||||
Db 199 GlyValIlyrHisValThrAsnAspCysSerAsnSerSerIleValIlyrGlyAlaAlaAsp 218
QY 304 ATGATCATGCAACACCCCGGGGTGCTGCTGCTGCTGCGGAGAACAACTTTCCCGCTGC 363
    |||||||
Db 219 MetIleMetHisIleAspProGlyCysValProCysValArgGlyAsnAsnSerSerArgCys 238
QY 364 TTGGTAGCGCTGACCCCGACGCTGCGAGCTAGGAAGCGACGCGCTCCACTAGGACATA 423
    |||||||
Db 239 TrpValAlaLeuThrProThrLeuAlaAlaArgAsnAlaSerValProThrThrIle 258
  
```

```

OY 424 CGAGCCACGTCGATTGCTGCTGGGGGGGCTGCTTCTGTTCGGTATGATGAGG 483
    |||||||
Db 259 ArgatGHISvalAspLeuValGlyAlaAlaAlaPheCysSerAlaMetGlyValGly 278
OY 484 GATCTCGGGATCTGTCTCTCTGTCCTCCAGCTGTTACACATCTCGCTCGGGGCA 543
    |||||||
Db 279 AspLeuCysGlySerValPheLeuValSerGlnLeuPheThrPheSerProArgArgHis 298
OY 544 GAGACGGTCAGAGCAATGCTCAATCTATCCCGGCGACATGAACAGAGTCACCGTATG 603
    |||||||
Db 259 GluThrValGlnAspCysAsnGlySerIleTyrProGlnHisLeuSerGlnHisArgMet 318
OY 604 GCTTGGAATGATGATGAACTGG 627
    |||||||
Db 319 AlaTrpAspMetMetMetAsnTrp 326

RESULT 9
O9QIT3 PRELIMINARY: PRT: 3010 AA.
AC 09QIT3:
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE Genome polyprotein [Contains: envelope glycoprotein E2 (GP68) (GP70)
    (NS1)].
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-Md-2;
RX MEDLINE=20013325; PubMed=10544098;
RA Nagayama K., Kurosaki M., Enomoto N., Maekawa S.Y., Miyasaka Y.,
    Tarawa J.I., Izumi N., Marumo F., Sato C.;
RT "Time-related changes in full-length hepatitis C virus and hepatitis
    activity."
RL Virology 263:244-253(1999).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-Md-4;
RA Nagayama K., Kurosaki M., Enomoto N., Maekawa S., Miyasaka Y.,
    Sakamoto N., Fukuma T., Tarawa J., Izumi N., Marumo F., Sato C.;
RL Submitted (JUL-1999) to the EMBL/GenBank/DBJ databases.
CC -1 SIMILARITY: TO HEPATITIS C VIRUS ENVELOPE GLYCOPROTEIN E1.
DR HSSP: P26663: 1UXP.
DR MEROPS: S29.001: -.
DR MEROPS: U39.001: -.
DR InterPro: IPR001410: DEAD.
DR InterPro: IPR002522: HCV_capsid.
DR InterPro: IPR002521: HCV_core.
DR InterPro: IPR002519: HCV_env.
DR InterPro: IPR002531: HCV_NS1.
DR InterPro: IPR002518: HCV_NS2.
DR InterPro: IPR004109: HCV_NS3.
DR InterPro: IPR000745: HCV_NS4.
DR InterPro: IPR001490: HCV_NS4b.
DR InterPro: IPR002868: HCV_NS5a.
DR InterPro: IPR002166: HCV_NS5b.
DR InterPro: IPR002166: HCV_NS5c.
DR Pfam: PF01543: HCV_capsid; 1.
DR Pfam: PF01542: HCV_core; 1.
DR Pfam: PF01539: HCV_env; 1.
DR Pfam: PF01560: HCV_NS1; 1.
DR Pfam: PF01538: HCV_NS2; 1.
DR Pfam: PF02907: HCV_NS3; 1.
DR Pfam: PF01006: HCV_NS4a; 1.
DR Pfam: PF01001: HCV_NS4b; 1.
DR Pfam: PF01506: HCV_NS5a; 1.
DR Pfam: PF00998: HCV_RdRp; 1.
DR ProDom: PD186062: HCV_NS1; 1.
KW ATP-binding; Coat protein; Envelope protein; Glycoprotein; Helicase;

```

```

KM Nonstructural protein; Polyprotein; RNA-directed RNA polymerase;
KW Transmembrane.
SQ SEQUENCE 3010 AA; 327087 MW; 2EF78321686D4002 CRC64;

Alignment Scores:
Pred. No.: 2,39e-96 Length: 3010
Score: 1086.00 Matches: 200
Percent Similarity: 99.04% Conservative: 6
Best Local Similarity: 96.15% Mismatches: 2
Query Match: 94.03% Indels: 0
DB: 12 Gaps: 0

US-09-899-303a-7 (1-633) x O9QIT3 (1-3010)
OY 4 TTGGTAAGATCATGATGATACCTTACGTCGAGGCTTCCGCGACCTGATGGGATACCTCG 63
    |||||||
Db 119 LeuGlyIysValIleAspThrIleuThrGlySerPheAlaAspLeuMetGlyTyrIlePro 138
OY 64 CTGTGCGGGCCCCCTAGGGGGTGTGCGCAGAGCCCTGGGCGATGGCGGTCTGTG 123
    |||||||
Db 139 LeuValGlyAlaProLeuGlyValAlaAlaAlaArgAlaLeuAlaHisGlyValAlaGlyAla 158
OY 124 GAAGACGGCGTGAATATGCAACAGAGGAATTTGCTGCTGCTTTCTATGCTCTGC 183
    |||||||
Db 159 GluAspGlyValAsnGlyAlaThrGlnAsnLeuProGlyCysSerPheSerIlePheLeu 178
OY 184 TTGGCTTACGTCTGCTGTCGATTCGACATTCGACCTTCCGCTTATGAGTGGCAAGCTGCC 243
    |||||||
Db 179 LeuAlaIleLeuSerCysLeuThrIleProAlaSerAlaTyrGluValAlaGlnValSer 198
OY 244 GGGATGACCATGTCACAGACAGACTGCTCAACTCAAGCATTTGTATGAGCGACGGAC 303
    |||||||
Db 199 GlyIleTyrHisValThrAsnAspCysSerAsnSerIleValTyrGluAlaAlaAsp 218
OY 304 AATGATGACGACACCCCGGGGTGGCTGCGCTGGCGGGAACAACCTTCCCGCTGC 363
    |||||||
Db 219 ValIleMetHisThrProGlyCysValProCysValArgGlnSerAsnSerIleGlyCys 238
OY 364 TTGGTAGCGCTACCCCGCAGCGCTCGACGTAGGAAGCCAGCGCTCCGACATGACGACATA 423
    |||||||
Db 239 TrpValAlaIleuThrProThrIleuAlaAlaArgAlaAlaSerValProThrThrThrIle 258
OY 424 CGAGCCACGTCGATTGCTGCTGGGGGGGCTGCTTCTGTTCGGTATGATGAGG 483
    |||||||
Db 259 ArgatGHISvalAspLeuValGlyThrAlaAlaAlaPheCysSerAlaMetGlyValGly 278
OY 484 GATCTCGGGATCTGTCTCTCTGTCCTCCAGCTGTTACACATCTCGCTCGGGGCA 543
    |||||||
Db 279 AspLeuCysGlySerValPheLeuValSerGlnLeuPheThrPheSerProArgArgHis 298
OY 544 GAGACGGTCAGAGCAATGCTCAATCTATCCCGGCGACATGAACAGAGTCACCGTATG 603
    |||||||
Db 259 GluThrValGlnAspCysAsnGlySerIleTyrProGlnHisLeuSerGlnHisArgMet 318
OY 604 GCTTGGAATGATGATGAACTGG 627
    |||||||
Db 319 AlaTrpAspMetMetMetAsnTrp 326

RESULT 10
O81760 PRELIMINARY: PRT: 3010 AA.
AC 081760:
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE Genome polyprotein [Contains: envelope glycoprotein E2 (GP68) (GP70)
    (NS1)].
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.

```



Alignment Scores:

Pred. No.:	Score:	Percent Similarity:	Best Local Similarity:	Query Match:	Length:	Matches:	Conservative:	Mismatches:	Indels:	Gaps:
2.98e-96	1085.00	98.56%	96.15%	93.94%	3010	200	5	3	0	0

US-09-899-303A-7 (1-633) x 0901Y4 (1-3010)

```

QY 4 TTGGGTAAGGTATGATGATCCCTTACGTGCGGCTTCCGCGACCTCATGGGGTACATTCCG 63
DB 119 LeuGlyLysValIleAspThrLeuThrCysGlyPheAlaAspLeuMetGlyTrpIlePro 138
QY 64 CTCGTGCGGCGCCCTAGCGGGTCTGCCAGAGCCCTGCGCATGGCTCCGGGTCTTG 123
DB 139 LeuValGlyAlaProLeuGlyGlyAlaAlaArgAlaLeuAlaHisGlyValAlaArgValLeu 158
QY 124 GAAGACGGCGTAACTATGCAACAGGGAATTGGCTGGTCTTTCTCTATCTTCTC 183
DB 159 GluAspGlyValAsnTyrlaThrGlyAsnLeuProGlyCysSerPheSerIlePheLeu 178
QY 184 TTGGGTTTACGTGCTGCTGTCACATTCAGCTTCCGTTAGAGTGGCGCAACGTCTCC 243
DB 179 LeuAlaLeuLeuSerCysLeuThrIleProAlaSerAlaTyrlaValAlaArgAsnValSer 198
QY 244 GGGATGTACCATGTATGCAAGACAGTCTCCAACTCAAGCATGTGTATGAGCGACGAC 303
DB 199 GlyIleTyrlaHisValIleThrAsnAspCysSerAsnSerIleValIleTyrlaAlaAlaAsp 218
QY 304 ATGATCATGCAACACCCCGGGTGGCTGCTGCTGCGGAGAACACTCTTCCGCTGC 363
DB 219 ValIleMetHisIleThrProIleCysValProCysValAlaGlyIleGlyAsnSerIleArgCys 238
QY 364 TGGGTAGCGCTCACCCCGCGCTCCAGCTAGAGACCGACGCGCTCCGCTCCAGACATA 423
DB 239 TrpValAlaLeuThrProThrIleAlaAlaArgAlaHisSerValProThrThrIle 258
QY 424 CGAGCGACAGTGGATTCCTGCTGCGGGGCGCTTCTGCTGCTCCGCTTACAGTGGG 483
DB 259 ArgArgHisValAlaAspLeuLeuValGlyThrAlaAlaPheCysSerAlaMetTyrlaGly 278
QY 484 GATCTCTGGGATGTGTCTTCTCTGCTGCTCCAGCTTTCACATCTCGCTCGCGGACAT 543
DB 279 AspLeuCysGlySerIlePheLeuValSerGlnLeuPheThrPheSerProArgHis 298
QY 544 GAGACGGTGCAGACTGCAATTGCTCAATCTATCCCGGCACATACAGTACAGCTAG 603
DB 299 GluThrValGlnAspCysAsnCysSerIleTyrlaProGlyHisValSerGlyHisArgMet 318
QY 604 GCTTGGGATATGATGATGACTGG 627
DB 319 AlaTrpAspMetMetAsnTrp 326
RESULT 12
0901Y4 PRELIMINARY: PRT; 3010 AA.
AC 0901Y4
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE Genome polyprotein [Contains: envelope glycoprotein E2 (GP68) (GP70) (NS1)].
OS Hepatitis C virus.
OC Viruses: ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RX MEDLINE=20013325; PubMed=10544098;
RA Nagayama K., Kurosaki M., Enomoto N., Maekawa S.Y., Miyasaka Y.,
Tazawa J.I., Izumi N., Marumo F., Sato C.;

```

"Time-related changes in full-length hepatitis C virus and hepatitis activity.";  
 RL Virology 263:244-253(1999).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-MD4-1;  
 RA Nagayama K., Kurosaki M., Enomoto N., Maekawa S., Miyasaka Y.,  
 Sakamoto N., Fukuma T., Tazawa J., Izumi N., Marumo F., Sato C.;  
 RL Submitted (JUL-1999) to the EMBL/Genbank/DBJ databases.  
 CC -1- SIMILARITY: TO HEPATITIS C VIRUS ENVELOPE GLYCOPROTEIN E1.  
 DR EMBL: AF165051; AAD56186.1; -.  
 DR HSRP: P26663; LXP.  
 DR MEROPS: S29.001; -.  
 DR MEROPS: U39.001; -.  
 DR InterPro: IPR001410; DEAD.  
 DR InterPro: IPR002522; HCV\_capsid.  
 DR InterPro: IPR002521; HCV\_core.  
 DR InterPro: IPR002519; HCV\_env.  
 DR InterPro: IPR002531; HCV\_NS1.  
 DR InterPro: IPR002518; HCV\_NS2.  
 DR InterPro: IPR004109; HCV\_NS3.  
 DR InterPro: IPR000745; HCV\_NS4a.  
 DR InterPro: IPR001490; HCV\_NS4b.  
 DR InterPro: IPR002868; HCV\_NS5a.  
 DR InterPro: IPR002166; HCV\_RdRP.  
 DR Pfam: PF01543; HCV\_capsid; 1.  
 DR Pfam: PF01542; HCV\_core; 1.  
 DR Pfam: PF01539; HCV\_env; 1.  
 DR Pfam: PF01560; HCV\_NS1; 1.  
 DR Pfam: PF01538; HCV\_NS2; 1.  
 DR Pfam: PF02907; HCV\_NS3; 1.  
 DR Pfam: PF01006; HCV\_NS4a; 1.  
 DR Pfam: PF01001; HCV\_NS4b; 1.  
 DR Pfam: PF01506; HCV\_NS5a; 1.  
 DR Pfam: PF00998; HCV\_RdRP; 1.  
 DR ProDom: PD186062; HCV\_NS1; 1.  
 KW ATP-binding; Coat protein; Envelope protein; Glycoprotein; Helicase;  
 KW Nonstructural protein; Polyprotein; RNA-directed RNA polymerase;  
 KW Transmembrane.  
 SQ SEQUENCE 3010 AA; 326830 MW; F6559AB2CFC53CB2 CRC64;

Alignment Scores:

Pred. No.:	Score:	Percent Similarity:	Best Local Similarity:	Query Match:	Length:	Matches:	Conservative:	Mismatches:	Indels:	Gaps:
2.98e-96	1085.00	99.04%	95.67%	93.94%	3010	199	7	2	0	0

US-09-899-303A-7 (1-633) x 0901Y4 (1-3010)

```

QY 4 TTGGGTAAGGTATGATGATCCCTTACGTGCGGCTTCCGCGACCTCATGGGGTACATTCCG 63
DB 119 LeuGlyLysValIleAspThrLeuThrCysGlyPheAlaAspLeuMetGlyTrpIlePro 138
QY 64 CTCGTGCGGCGCCCTAGCGGGTCTGCCAGAGCCCTGCGCATGGCTCCGGGTCTTG 123
DB 139 LeuValGlyAlaProLeuGlyGlyAlaAlaArgAlaLeuAlaHisGlyValAlaArgValLeu 158
QY 124 GAAGACGGCGTAACTATGCAACAGGGAATTGGCTGGTCTTTCTCTATCTTCTC 183
DB 159 GluAspGlyValAsnTyrlaThrGlyAsnLeuProGlyCysSerPheSerIlePheLeu 178
QY 184 TTGGGTTTACGTGCTGCTGTCACATTCAGCTTCCGTTAGAGTGGCGCAACGTCTCC 243
DB 179 LeuAlaLeuLeuSerCysLeuThrIleProAlaSerAlaTyrlaValAlaArgAsnValSer 198
QY 244 GGGATGTACCATGTATGCAAGACAGTCTCCAACTCAAGCATGTGTATGAGCGACGAC 303
DB 199 GlyIleTyrlaHisValIleThrAsnAspCysSerAsnSerIleValIleTyrlaAlaAlaAsp 218
QY 304 ATGATCATGCAACACCCCGGGTGGCTGCTGCTGCGGAGAACACTCTTCCGCTGC 363

```

```

Db 219 ValIleMetHisThrProGlyCysValaProCysValaArgIleSerAsnSerSerArgCys 238
QY 364 TGGGTAGCGCTCAACCCCGAGCTGCAGTACAGGAGCCAGCGTCCCTACGACATATA 423
Db 239 TrpValaIleuThrProThrIleuAlaIleArgAsnIleSerIleProThrThrIle 258
QY 424 CGAGCCGACGCTGCATTTGCTGTTGGGGCGGCGCTTTCCTTCCTGCTATGACGTGGG 483
Db 259 ArgArgHisValaIleuValaIleGlyThrAlaIlePheCysSerAlaMetIleValaIle 278
QY 484 GATCTGCGGGGATCTGCTCTCCGTCGTCCTCCAGCTGTTCCACATCTGCGCTCGCGGCAT 543
Db 279 AspIleuCysGlySerValaPheIleuIleSerGlnIleuPheSerProArgArgHis 298
QY 544 GAGACGGTGCAGGACGTGCATTTGCTCAATCTATCCCGGCACATTAACAGGTACCGGTATG 603
Db 299 GlnThrValaGlnAspCysAsnCysSerIleIleTyrProGlyHisValaSerGlyHisArgMet 318
QY 604 GCTTGGGATATGATGATGAACGTG 627
Db 319 AlaTrpAspMetMetMetAsnTrp 326

RESULT 13
ID 081776 PRELIMINARY; PRT; 441 AA.
AC 081776;
DT 01-NOV-1996 (TREMBLER. 01, Created)
DT 01-NOV-1996 (TREMBLER. 01, Last sequence update)
DT 01-DEC-2001 (TREMBLER. 19, Last annotation update)
DE Genome polyprotein [contains: envelope glycoprotein E2 (GP68) (GP70)
DE (NS1)] (Fragment).
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
NC NCBL_TaxID-11103;
RN [1]
RP SEQUENCE OF 325-441 FROM N.A.
RC STRAIN-JH/JAPAN/J7 AND JH/JAPAN/J1;
RX MEDLINE-91007289; PubMed-2170237;
RA Takeuchi K., Boonmar S., Kubo Y., Katayama T., Harada H.,
RA Ohbayashi A., Choo Q., Kuo G., Houghton M., Saito I., Miyamura T.;
RT "Hepatitis C viral cDNA clones isolated from a healthy carrier donor
RT implicated in post-transfusion non-A/non-B hepatitis.";
RL Gene 91:287-291(1990).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-JH/JAPAN/J7 AND JH/JAPAN/J1;
RX MEDLINE-90356432; PubMed-2117749;
RA Takeuchi K., Kubo Y., Boonmar S., Watanabe Y., Katayama T., Choo Q.,
RA Kuo G., Houghton M., Saito I., Miyamura T.;
RT "Nucleotide sequence of core and envelope genes of the hepatitis C
RT virus genome derived directly from human healthy carriers.";
RL Nucleic Acids Res. 18:4626-4626(1990).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN-JH/JAPAN/J7 AND JH/JAPAN/J1;
RX MEDLINE-94270990; PubMed-7545932;
RA Hotta H., Doi H., Hayashi T., Furukawa M., Soemarto W., Mizokami M.,
RA Ohba K., Homma M.;
RT "Analysis of the core and E1 envelope region sequences of a novel
RT variant of hepatitis C virus obtained in Indonesia.";
RL Arch. Virol. 136:53-62(1994).
CC -1- SIMILARITY: TO HEPATITIS C VIRUS ENVELOPE GLYCOPROTEIN E1.
DR EMBL; D00574; BAA00452.1; -.
DR InterPro: IPRO02522; HCV_capsid.
DR InterPro: IPRO02521; HCV_core.
DR InterPro: IPRO02519; HCV_env.
DR InterPro: IPRO02531; HCV_NSI.
DR Pfam; PF01543; HCV_capsid; 1.
DR Pfam; PF01542; HCV_core; 1.
DR Pfam; PF01539; HCV_env; 1.
DR Pfam; PF01560; HCV_NSI; 1.
DR ProDom; PD186062; HCV_NSI; 1.

```

```

KW Coat protein; Envelope protein; Glycoprotein; Nonstructural protein;
KW Polyprotein; Transmembrane.
FT NON_TER 441 441
SQ SEQUENCE 441 AA: 47875 MW: 78025F07957EFA1E CRC64:

Alignment Scores:
Pred. No.: 2,86e-96 Length: 441
Score: 1084.00 Matches: 199
Percent Similarity: 98.56% Conservative: 6
Best Local Similarity: 95.67% Mismatches: 3
Query Match: 93.85% Indels: 0
DB: 12 Gaps: 0

US-09-899-303a-7 (1-633) x 081776 (1-441)
QY 4 TTGGGTAGGTCATTCGATACCTTAGTGCGGCTTGCGCCAGCTCATGGGTACATTCCG 63
Db 119 LeuGlyValaIleAspThrIleuThrCysGlyPheAlaAspIleuMetGlyTyrIlePro 138
QY 64 CTCGTGCGCGCCCGCTAGGGGTGCTGCCAGAGCCCTGGCGCATGGCGGTTCG 123
Db 139 LeuValaGlyAlaProIleuGlyGlyAlaIleArgAlaIleuAlaHisGlyValaArgValleu 158
QY 124 GAAAGCGCGCTGAACATATGCACAGGGAATTTGCTGCTTCTCTATCTTCCTC 183
Db 159 GluAspGlyValaAsnTyrAlaThrGlyAsnIleuProGlyCysSerPheSerIlePheIleu 178
QY 184 TTGGCTTTACTGTCTGCTGACCATTCACGCTTCGCTTATAGTGCGCAACGTGTC 243
Db 179 LeuAlaIleuSerCysIleuThrIleProAlaSerAlaTyrIleValaArgAsnValaSer 198
QY 244 GGGATGACCATGTCAGAGACGACTGCTCCAACTCAAGCATGTTGATAGGAGCGGAC 303
Db 199 GlyIleTyrHisValaIleAsnAspCysSerAsnSerSerIleValaIleGlyAlaIleAsp 218
QY 304 ATGATCATGCACACCCCGGGTGCCTGCTGCGCTTCGGAGAACATCTTCCGCTCG 363
Db 219 ValIleMetHisThrProGlyCysValaProCysValaArgIleSerAsnSerSerArgCys 238
QY 364 TGGGTAGCGCTCAACCCCGAGCTGCAGTACAGGAGCCAGCGTCCCTACGACATATA 423
Db 239 TrpValaIleuThrProThrIleuAlaIleArgAsnIleSerIleProThrThrIle 258
QY 424 CGAGCCGACGCTGCATTTGCTGTTGGGGCGGCGCTTTCCTTCCTGCTATGACGTGGG 483
Db 259 ArgArgHisValaIleuValaIleGlyThrAlaIlePheCysSerAlaMetIleValaIle 278
QY 484 GATCTGCGGGATCTGCTCTCCGTCGTCCTCCAGCTGTTCCACATCTGCGCTCGCGGCAT 543
Db 279 AspIleuCysGlySerValaPheIleuIleSerGlnIleuPheSerProArgArgHis 298
QY 544 GAGACGGTGCAGGACGTGCATTTGCTCAATCTATCCCGGCACATTAACAGGTACCGGTATG 603
Db 299 GlnThrValaGlnAspCysAsnCysSerIleIleTyrProGlyHisValaSerGlyHisArgMet 318
QY 604 GCTTGGGATATGATGATGAACGTG 627
Db 319 AlaTrpAspMetMetMetAsnTrp 326

RESULT 14
ID 068520 PRELIMINARY; PRT; 802 AA.
AC 068520;
DT 01-NOV-1996 (TREMBLER. 01, Created)
DT 01-NOV-1996 (TREMBLER. 01, Last sequence update)
DT 01-DEC-2001 (TREMBLER. 19, Last annotation update)
DE Genome polyprotein [contains: envelope glycoprotein E2 (GP68) (GP70)
DE (NS1)] (Fragment).
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
NC NCBL_TaxID-11103;
RN [1]

```

```

RP SEQUENCE FROM N.A.
RC STRAIN-HTX.
RA Mueller H.W., Selzer S., Heller A.E., Pfaff E., Theilmann L.;
RT "Recurrent hepatitis C virus infection after liver transplantation in
RT a patient negative for viral proteins: A possible role of immune
RT tolerance.";
RU Submitted (Jan-1996) to the EMBL/Genbank/DBJ databases.
CC -1- SIMILARITY: TO HEPATITIS C VIRUS ENVELOPE GLYCOPROTEIN E1.
DR EMBL: U45463; AAA86919.1; -.
DR MEROPS: U39.001; -.
DR InterPro: IPR002522; HCV_capsid.
DR InterPro: IPR002521; HCV_core.
DR InterPro: IPR002519; HCV_env.
DR InterPro: IPR002531; HCV_NS1.
DR InterPro: IPR002518; HCV_NS2.
DR Pfam: PF01543; HCV_capsid; 1.
DR Pfam: PF01542; HCV_core; 1.
DR Pfam: PF01539; HCV_env; 1.
DR Pfam: PF01560; HCV_NS1; 1.
DR Pfam: PF01538; HCV_NS2; 1.
DR ProDom: PD186062; HCV_NS1; 1.
DR CoaT protein: Envelope protein; glycoprotein; Nonstructural protein;
KW Polypeptide; Transmembrane.
FT NON_TER
SQ SEQUENCE 802 AA; 87900 MW; F17246CAAAA476 CRC64;

Alignment Scores:
Pred. No.: 3,11e-96 Length: 802
Score: 1084.00 Matches: 201
Percent Similarity: 99.04% Conservative: 5
Best Local Similarity: 96.63% Mismatches: 2
Query Match: 93.85% Indels: 0
DB: 12 Gaps: 0

US-09-899-303A-7 (1-633) x Q68520 (1-802)
OY 4 TTGGGTAAGGCTATGATACCTTACGTCGGCTTCGCCGACCTCATGSGGTACATTCCG 63
DB 119 LeuGlyLysValIleAspThrLeuThrCysGlyPheAlaAspLeuMetGlyTyrIlePro 138
OY 64 CTGCTGCGGGCCCCCTAGGGGGTGGTCCAGAGCCCTGGCCGATGGCTCGGGGTTCG 123
DB 139 LeuValGlyAlaProLeuGlyGlyAlaAlaArgAlaLeuAlaHisGlyAlaArgValIleu 158
OY 124 GAAGACGGCGTAACATAAGCAAGGAAATTGCTGCTTTCTATCTTCTCTC 183
DB 159 GluAspGlyValAsnTyrAlaThrGlyAsnLeuProGlySerPheSerIlePheLeu 178
OY 184 TTGGGTTTACGTCGTCGTGACCAATTCACGTCCTCCGCTATGAGTGGCAACGTCTCC 243
DB 179 LeuAlaLeuLeuSerCysLeuThrIleProAlaSerAlaTyrGluValArgAsnValSer 198
OY 244 GGGATGACCATGATCAGAGACAGTGTCTCAACTCAAGCATGTGTATGAGCAGCGGAC 303
DB 199 GlyValTyrHisValIleThrAsnAspCysSerAsnSerIleValTyrIleAlaAlaAsp 218
OY 304 ATGATCATGACACACCCCGGGTGGTCCGCTTCGGGAGAACAACATCTTCCGCTGC 363
DB 219 MetIleMetHisThrProGlyValProGlyValArgGluAsnAsnSerSerArgCys 238
OY 364 TGGGTAGCGCTCACCCGCCAGCTCCGAGTAGAAGACCCAGCGTCCCACTACGACAAATA 423
DB 239 TrpValAlaLeuThrProThrLeuAlaAlaArgSerAlaAsnValSerThrThrIle 258
OY 424 CGACCCACGATGATTTGCTGCTGGGGCGGTCTTTGTTCGCTATGATGAGTGGGG 483
DB 259 ArgAlaGHisValAspLeuLeuValGlyAlaAlaAlaPheCysSerAlaMetTyrValGly 278
OY 484 GATCTCTGGGATCTGTCTTCTCTGCTTCGACGCTGTTACCATCTCGCCTCGCGGCAT 543
DB 279 AspLeuCysGlySerValPheLeuValSerGlnLeuPheThrPheSerProArgAlaGHis 298
OY 544 GAGACGGTCCAGGACTGCAATTGCTCAATCTATCCCGGCCACATAAGAGGTACCCGTATG 603

```

```

DB 299 GluThrValGlnAspCysAsnCysSerIleTyrProGlyHisValSerGlyHisArgMet 318
OY 604 GCTTGCGATGATGATGATGACTGG 627
DB 319 AlaTrpAspMetMetMetAsnTrp 326

RESULT 15
ID Q9J3H7 PRELIMINARY; PRT; 3010 AA.
AC Q9J3H7;
DT 01-OCT-2000 (TREMBLrel. 15, Created)
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE Genome polypeptide [Contains: envelope glycoprotein E2 (GP68) (GP70)
DE (NS1)].
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-MD15;
RA Nagayama K., Kurosaki M., Enomoto N., Miyasaka Y., Marumo F., Sato C.;
RT "Characteristics of hepatitis C viral genome associated with disease
RT progression.";
RT Submitted (NOV-1999) to the EMBL/Genbank/DBJ databases.
CC -1- SIMILARITY: TO HEPATITIS C VIRUS ENVELOPE GLYCOPROTEIN E1.
DR EMBL: AF207756; AAF65946.1; -.
DR HSRP: p26663; 1JXP.
DR MEROPS: S29.001; -.
DR MEROPS: U39.001; -.
DR InterPro: IPR000345; Cytochrome_b10d.
DR InterPro: IPR001410; DEAD.
DR InterPro: IPR002522; HCV_capsid.
DR InterPro: IPR002521; HCV_core.
DR InterPro: IPR002519; HCV_env.
DR InterPro: IPR002531; HCV_NS1.
DR InterPro: IPR002518; HCV_NS2.
DR InterPro: IPR004109; HCV_NS3.
DR InterPro: IPR000745; HCV_NS4a.
DR InterPro: IPR001490; HCV_NS4b.
DR InterPro: IPR002868; HCV_NS5a.
DR InterPro: IPR002166; HCV_RdRp.
DR Pfam: PF01543; HCV_capsid; 1.
DR Pfam: PF01542; HCV_core; 1.
DR Pfam: PF01539; HCV_env; 1.
DR Pfam: PF01560; HCV_NS1; 1.
DR Pfam: PF01538; HCV_NS2; 1.
DR Pfam: PF02907; HCV_NS3; 1.
DR Pfam: PF02907; HCV_NS4; 1.
DR Pfam: PF01006; HCV_NS4a; 1.
DR Pfam: PF01001; HCV_NS4b; 1.
DR Pfam: PF01506; HCV_NS5a; 1.
DR Pfam: PF00998; HCV_RdRp; 1.
DR ProDom: PD186062; HCV_NS1; 1.
DR PROSITE: PS00190; CYTOCHROME_C; UNKNOWN_1.
KW CoaT protein; Envelope protein; glycoprotein; Nonstructural protein;
KW Polypeptide; RNA-directed RNA polymerase; Transmembrane.
SQ SEQUENCE 3010 AA; 327365 MW; D8653F7317FFA106 CRC64;

Alignment Scores:
Pred. No.: 3.72e-96 Length: 3010
Score: 1084.00 Matches: 200
Percent Similarity: 97.60% Conservative: 3
Best Local Similarity: 96.15% Mismatches: 5
Query Match: 93.85% Indels: 0
DB: 12 Gaps: 0

US-09-899-303A-7 (1-633) x Q9J3H7 (1-3010)
OY 4 TTGGTAAAGGCTATGATACCTTACGTCGGCTTCGCCGACCTCATGSGGTACATTCCG 63
DB 119 LeuGlyLysValIleAspThrLeuThrCysGlyPheAlaAspLeuMetGlyTyrIlePro 138

```

```
QY 64 CTGCTGGGCCCCCTAGGGGGTGTCCAGACCCCTGGCGCATGGCGTCCGGGTTCTG 123
    |||||
Db 139 LeuValGlyAlaProLeuGlyGlyAlaAlaArgAlaLeuAlaHisGlyValArgValLeu 158
    |||||
QY 124 GAGACGGCGCTGAACATGATGACAGGAAATTTGGCTGTCTCTATCTTCTC 183
    |||||
Db 159 GluAspGlyValAsnGlyAlaThrGlyAsnLeuProGlyCysSerPheSerLeuPheLeu 178
    |||||
QY 184 TTGGCTTTACTGTCCTGCTGACCATTTCCAGCTTCCGCTATGAGTGGCCAACTGTC 243
    |||||
Db 179 LeuAlaLeuLeuSerCysLeuThrProValSerIleArgIleValAlaArgAsnValSer 198
    |||||
QY 244 GGGATGATACCATGTCACGAAAGACTGCTCCAACTCAAGCATTTGTATGAGGACGGAC 303
    |||||
Db 199 GlyValTyrHisValThrAsnAspCysSerAsnSerSerIleValTyrGluAlaAlaAsp 218
    |||||
QY 304 ATGATCATGACACACCCCGGGGTGGTGGCTTGGGAGAGAACACTCTCCCGCTGC 363
    |||||
Db 219 MetIleMetHisAlaProGlyCysValProCysValArgGluAsnAsnSerSerArgCys 238
    |||||
QY 364 TGGGTAGCGCTCACCCCGACGCTCGCAGTAGGAACGCCACCGTCCCGCTACGACAATA 423
    |||||
Db 239 TrpValAlaLeuThrProThrLeuAlaAlaArgAsnIleSerIleProThrThrIle 258
    |||||
QY 424 GACGCCACGCTCATTTGCTGTTGGGGCGGCTGCTTCTGTCCGCTATGATGAGTGGG 483
    |||||
Db 259 ArgArgHisValAspLeuLeuValGlyAlaAlaAlaPheCysSerAlaMetTyrValGly 278
    |||||
QY 484 GATCTCTGGCGGATCTGTCTCTCCGCTGCCAGCTGTTCACCATCTGCGCTCGCGGAT 543
    |||||
Db 279 AspLeuCysGlySerValPheLeuValSerGlnLeuPheThrPheSerProArgArgHis 298
    |||||
QY 544 GAGACGGTGCAGACATGCAATTTGCTCATCTATCCCGCCACATACAGGTACACCGTATG 603
    |||||
Db 299 GluThrValGlnAspCysAsnCysSerIleTyrProGlyHisValThrGlyHisArgMet 318
    |||||
QY 604 GCTTGGGATATGATGATGAACCTGG 627
    |||||
Db 319 AlaTrpAspMetMetMetAsnTrp 326
    |||||
```

Search completed: February 12, 2003, 09:01:01  
Job time : 79 secs

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.3  
Copyright (c) 1993 - 2003 CompuGen Ltd.

# OM nucleic - nucleic search, using sw model

Run on: February 18, 2003, 04:57:02 ; Search time 2140 Seconds

(without alignments)  
8608.447 Million cell updates/sec

Title: US-09-899-303a-7

Perfect score: 633  
Sequence: 1 ATGTGGGTAAAGTCATCGA.....TGATGATGAACGTGTAATAG 633

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

## Database :

1: gb\_ba:\*  
2: gb\_hcg:\*  
3: gb\_in:\*  
4: gb\_om:\*  
5: gb\_ov:\*  
6: gb\_pat:\*  
7: gb\_ph:\*  
8: gb\_pl:\*  
9: gb\_pr:\*  
10: gb\_ro:\*  
11: gb\_sts:\*  
12: gb\_sy:\*  
13: gb\_un:\*  
14: gb\_vl:\*  
15: em\_ba:\*  
16: em\_fun:\*  
17: em\_hum:\*  
18: em\_in:\*  
19: em\_mu:\*  
20: em\_om:\*  
21: em\_or:\*  
22: em\_ov:\*  
23: em\_pat:\*  
24: em\_ph:\*  
25: em\_pl:\*  
26: em\_ro:\*  
27: em\_sts:\*  
28: em\_un:\*  
29: em\_vl:\*  
30: em\_hcg\_hum:\*  
31: em\_hcg\_inv:\*  
32: em\_hcg\_other:\*  
33: em\_hcg\_mus:\*  
34: em\_hcg\_pln:\*  
35: em\_hcg\_rod:\*  
36: em\_hcg\_man:\*  
37: em\_hcg\_vrt:\*  
38: em\_sy:\*  
39: em\_hcgo\_hum:\*  
40: em\_hcgo\_mus:\*  
41: em\_hcgo\_other:\*

Score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	633	100.0	633	6 A48669	A48669 Sequence 7
2	633	100.0	633	6 A452756	A452756 Sequence
3	633	100.0	633	6 A48675	A48675 Sequence 13
4	612.8	96.8	636	6 A48675	A48675 Sequence 13
5	612.8	96.8	636	6 A48675	A48675 Sequence 13
6	612.8	96.8	636	6 A48675	A48675 Sequence 13
7	608.8	96.2	795	6 A48667	A48667 Sequence 5
8	608.8	96.2	795	6 A48667	A48667 Sequence 5
9	608.8	96.2	795	6 A48667	A48667 Sequence 5
10	605.8	95.7	2082	6 A48709	A48709 Sequence 47
11	605.8	95.7	2082	6 A48709	A48709 Sequence 47
12	605.8	95.7	2082	6 A48709	A48709 Sequence 47
13	605.8	95.7	2082	6 A48709	A48709 Sequence 47
14	605.8	95.7	2082	6 A48709	A48709 Sequence 47
15	605.8	95.7	2082	6 A48709	A48709 Sequence 47
16	574.2	90.7	9379	14 A48709	A48709 Sequence 47
17	569.4	90.0	9386	14 A48709	A48709 Sequence 47
18	567.8	89.7	9379	14 A48709	A48709 Sequence 47
19	567.8	89.7	9379	14 A48709	A48709 Sequence 47
20	567.8	89.7	9386	14 A48709	A48709 Sequence 47
21	566.2	89.4	9410	14 A48709	A48709 Sequence 47
22	564.6	89.2	1880	14 A48709	A48709 Sequence 47
23	564.6	89.2	2540	6 E04805	E04805 CDNA encod1
24	564.6	89.2	2540	6 E04805	E04805 CDNA encod1
25	564.6	89.2	2540	6 E04805	E04805 CDNA encod1
26	564.6	89.2	2540	6 E04805	E04805 CDNA encod1
27	564.6	89.2	2540	6 E04805	E04805 CDNA encod1
28	564.6	89.2	2540	6 E04805	E04805 CDNA encod1
29	564.6	89.2	2540	6 E04805	E04805 CDNA encod1
30	564.6	89.2	2540	6 E04805	E04805 CDNA encod1
31	564.6	89.2	2540	6 E04805	E04805 CDNA encod1
32	564.6	89.2	2540	6 E04805	E04805 CDNA encod1
33	564.6	89.2	2540	6 E04805	E04805 CDNA encod1
34	564.6	89.2	2540	6 E04805	E04805 CDNA encod1
35	564.6	89.2	2540	6 E04805	E04805 CDNA encod1
36	564.6	89.2	2540	6 E04805	E04805 CDNA encod1
37	564.6	89.2	2540	6 E04805	E04805 CDNA encod1
38	564.6	89.2	2540	6 E04805	E04805 CDNA encod1
39	564.6	89.2	2540	6 E04805	E04805 CDNA encod1
40	564.6	89.2	2540	6 E04805	E04805 CDNA encod1
41	564.6	89.2	2540	6 E04805	E04805 CDNA encod1
42	564.6	89.2	2540	6 E04805	E04805 CDNA encod1
43	564.6	89.2	2540	6 E04805	E04805 CDNA encod1
44	564.6	89.2	2540	6 E04805	E04805 CDNA encod1
45	564.6	89.2	2540	6 E04805	E04805 CDNA encod1

## ALIGNMENTS

RESULT 1  
LOCUS A48669 633 bp DNA linear PAT 07-MAR-1997  
DEFINITION Sequence 7 from Patent WO9604385.  
ACCESSION A48669  
VERSION A48669.1 GI:2302382  
KEYWORDS  
SOURCE unidentified.  
ORGANISM unidentified.  
REFERENCE 1 (bases 1 to 633)  
AUTHORS Maertens G., Bosman F., De M.G. and Buysse M.  
TITLE PURIFIED HEPATITIS C VIRUS ENVELOPE PROTEINS FOR DIAGNOSTIC AND  
THERAPEUTIC USE  
JOURNAL Patent: WO 9604385-A 7 15-FEB-1996;

Pred. No. is the number of results predicted by chance to have a

COMMENT INNOGENETICS NV (BE)  
Other publication CA 2172273 960215  
Other publication AU 3382495 960304.  
FEATURES  
source  
1. 633  
Location/Qualifiers

CDS  
/organism="unidentified"  
/db\_xref="taxon:32644"  
1. 630  
/note="unnamed protein product"  
/codon\_start=1  
/protein\_id="CAA03130.1"  
/db\_xref="GI:2302383"  
translation="MLGKVIDPLTGFADIMGYIPLVGNPLGAGARALAHGVRLDGV  
VYATGNLPGCSFSEIFLALISCLTPASAYEVRVSGMYHVNCSSTIYEADIM  
IMHPPGCTPCVRENSSRCWALPTTLARASVPTTTRRVDLLVGAAPCSAMIV  
GDLGCVFLVSOLFETISPRRHETVODNCSTYPGHITGRHAMDMMNMW"  
1. 627  
mat\_peptide  
/product="unnamed"  
BASE COUNT 111 a 192 c 174 g 156 t  
ORIGIN

Query Match 100.0%; Score 633; DB 6; Length 633;  
Best Local Similarity 100.0%; Pred. No. 6.4e-144;  
Matches 633; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 ATGTTGGTAAGTCAATCATACCTTACGTGGGCTTCGGACCTCATGGGTACATT 60  
1 ATGTTGGTAAGTCAATCATACCTTACGTGGGCTTCGGACCTCATGGGTACATT 60  
61 CCGCTGTGGGGCCCCCTAGGGGGTGTGCGCAGAGCCCTGGCGCATGGCGGGTT 120  
61 CCGCTGTGGGGCCCCCTAGGGGGTGTGCGCAGAGCCCTGGCGCATGGCGGGTT 120  
121 CTGGAAGAGCGGGTGAATCATGCAACAGAGGAATTGCTGTGCTCTTCTATCTTC 180  
121 CTGGAAGAGCGGGTGAATCATGCAACAGAGGAATTGCTGTGCTCTTCTATCTTC 180  
121 CTGGAAGAGCGGGTGAATCATGCAACAGAGGAATTGCTGTGCTCTTCTATCTTC 180  
181 CTCTTGCTTTACTGCTGCTGCTGACCATTCAGCTTCGGCTTATAGGTGGCAACGTG 240  
181 CTCTTGCTTTACTGCTGCTGCTGACCATTCAGCTTCGGCTTATAGGTGGCAACGTG 240  
181 CTCTTGCTTTACTGCTGCTGCTGACCATTCAGCTTCGGCTTATAGGTGGCAACGTG 240  
241 TCCGGGATGTACCATGTGCAAGAGAGCTGCTCAACTCAAGCATTTGTATAGGACGG 300  
241 TCCGGGATGTACCATGTGCAAGAGAGCTGCTCAACTCAAGCATTTGTATAGGACGG 300  
301 GACATGATCATGCACACCCCGGGGTGCTGCTGCTGCGGTTGCGGAGAACAACTTCCCGC 360  
301 GACATGATCATGCACACCCCGGGGTGCTGCTGCTGCGGTTGCGGAGAACAACTTCCCGC 360  
361 TCGTGGGTAGCGCTACCCCGACGCTCGAGCTAGGAGACGCCAGCTCCCACTACGACA 420  
361 TCGTGGGTAGCGCTACCCCGACGCTCGAGCTAGGAGACGCCAGCTCCCACTACGACA 420  
421 ATACGACGCGACGTCGATTTGCTGCTGGGGGCGCTGCTTCTGTTCCGCTATGACGTG 480  
421 ATACGACGCGACGTCGATTTGCTGCTGGGGGCGCTGCTTCTGTTCCGCTATGACGTG 480  
421 ATACGACGCGACGTCGATTTGCTGCTGGGGGCGCTGCTTCTGTTCCGCTATGACGTG 480  
481 GGGGATCTCTGGGGATCTGCTCTCTGCTCCAGCTGTTCAACATCTCCGCTCCCGG 540  
481 GGGGATCTCTGGGGATCTGCTCTCTGCTCCAGCTGTTCAACATCTCCGCTCCCGG 540  
541 CATGAGACGCTGACGAGCTGCAATTCCTCAATCTATCCGCGCACATAACAGGTACCGGT 600  
541 CATGAGACGCTGACGAGCTGCAATTCCTCAATCTATCCGCGCACATAACAGGTACCGGT 600  
601 ATGGCTTGGGATATGATGATGAAGCTGTATAG 633  
601 ATGGCTTGGGATATGATGATGAAGCTGTATAG 633

RESULT 2  
ARI57326  
LOCUS ARI57326 633 bp DNA linear PAT 17-OCT-2001

DEFINITION Sequence 7 from patent US 6245503.  
ACCESSION ARI57326  
VERSION ARI57326.1 GI:16218259  
KEYWORDS  
SOURCE  
ORGANISM  
REFERENCE  
1 (bases 1 to 633)  
AUTHORS  
TITLE  
JOURNAL  
FEATURES  
source  
1. 633  
Location/Qualifiers  
BASE COUNT 111 a 192 c 174 g 156 t  
ORIGIN

Query Match 100.0%; Score 633; DB 6; Length 633;  
Best Local Similarity 100.0%; Pred. No. 6.4e-144;  
Matches 633; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 ATGTTGGTAAGTCAATCATACCTTACGTGGGCTTCGGACCTCATGGGTACATT 60  
1 ATGTTGGTAAGTCAATCATACCTTACGTGGGCTTCGGACCTCATGGGTACATT 60  
61 CCGCTGTGGGGCCCCCTAGGGGGTGTGCGCAGAGCCCTGGCGCATGGCGGGTT 120  
61 CCGCTGTGGGGCCCCCTAGGGGGTGTGCGCAGAGCCCTGGCGCATGGCGGGTT 120  
121 CTGGAAGAGCGGGTGAATCATGCAACAGAGGAATTGCTGTGCTCTTCTATCTTC 180  
121 CTGGAAGAGCGGGTGAATCATGCAACAGAGGAATTGCTGTGCTCTTCTATCTTC 180  
121 CTGGAAGAGCGGGTGAATCATGCAACAGAGGAATTGCTGTGCTCTTCTATCTTC 180  
181 CTCTTGCTTTACTGCTGCTGCTGACCATTCAGCTTCGGCTTATAGGTGGCAACGTG 240  
181 CTCTTGCTTTACTGCTGCTGCTGACCATTCAGCTTCGGCTTATAGGTGGCAACGTG 240  
181 CTCTTGCTTTACTGCTGCTGCTGACCATTCAGCTTCGGCTTATAGGTGGCAACGTG 240  
241 TCCGGGATGTACCATGTGCAAGAGAGCTGCTCAACTCAAGCATTTGTATAGGACGG 300  
241 TCCGGGATGTACCATGTGCAAGAGAGCTGCTCAACTCAAGCATTTGTATAGGACGG 300  
301 GACATGATCATGCACACCCCGGGGTGCTGCTGCTGCGGTTGCGGAGAACAACTTCCCGC 360  
301 GACATGATCATGCACACCCCGGGGTGCTGCTGCTGCGGTTGCGGAGAACAACTTCCCGC 360  
361 TCGTGGGTAGCGCTACCCCGACGCTCGAGCTAGGAGACGCCAGCTCCCACTACGACA 420  
361 TCGTGGGTAGCGCTACCCCGACGCTCGAGCTAGGAGACGCCAGCTCCCACTACGACA 420  
421 ATACGACGCGACGTCGATTTGCTGCTGGGGGCGCTGCTTCTGTTCCGCTATGACGTG 480  
421 ATACGACGCGACGTCGATTTGCTGCTGGGGGCGCTGCTTCTGTTCCGCTATGACGTG 480  
421 ATACGACGCGACGTCGATTTGCTGCTGGGGGCGCTGCTTCTGTTCCGCTATGACGTG 480  
481 GGGGATCTCTGGGGATCTGCTCTCTGCTCCAGCTGTTCAACATCTCCGCTCCCGG 540  
481 GGGGATCTCTGGGGATCTGCTCTCTGCTCCAGCTGTTCAACATCTCCGCTCCCGG 540  
541 CATGAGACGCTGACGAGCTGCAATTCCTCAATCTATCCGCGCACATAACAGGTACCGGT 600  
541 CATGAGACGCTGACGAGCTGCAATTCCTCAATCTATCCGCGCACATAACAGGTACCGGT 600  
601 ATGGCTTGGGATATGATGATGAAGCTGTATAG 633  
601 ATGGCTTGGGATATGATGATGAAGCTGTATAG 633

RESULT 3  
AX452756  
LOCUS AX452756 633 bp DNA linear PAT 06-JUL-2002  
DEFINITION Sequence 7 from Patent EP2013305  
ACCESSION AX452756  
VERSION AX452756.1 GI:24712441

KEYWORDS  
SOURCE Hepatitis C virus.  
ORGANISM Hepatitis C virus  
Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae; Hepacivirus.

REFERENCE  
1 Maertens,G., Bosman,F., de Martynoff,G. and Buysse,M.A.  
TITLE Recombinant vectors for producing hcv envelope proteins.  
JOURNAL Patent: EP 121315-A 7-05-JUN-2002  
Inogenetics N.V. (BE)

FEATURES  
source  
1. 633  
Location/Qualifiers  
/organism="Hepatitis C virus"  
/db\_xref="taxon:11103"  
1. .630  
CDS  
/note="unnamed protein product"  
/codon\_start=1  
/protein\_id="CAD38068.1"  
/db\_xref="GI:21712442"  
/translation="MLGRVLDITLCGFADLMGYIDLVGAPLGGAAALAHGVRLDGVNATGNLPGCSFSLFLALSLCTIPASAEVRNVSQMYHTNDSSNSIYEADIMHTPGCVPCVRENSSRCWVALPTPLAARNASVPTTIRRHVDLLVGAAPCSAMYGDIGSVFLVSQLFTISPRRHETVQDCNSIYPGHITGHRAMDMMMW"

mat\_peptide  
1. 627  
/product="unnamed"

BASE COUNT 111 a 192 c 174 g 156 t

ORIGIN

Query Match 100.0%; Score 633; DB 6; Length 633;  
Best Local Similarity 100.0%; Pred. No. 6.4e-144; Mismatches 0; Indels 0; Gaps 0;  
Matches 633; Conservative 0;

Db 1 ATGTTGGGTAGATCATCATCCTTACCTAGTGGGCTTCCGACCTCATGGGGTACATT 60  
1 ATGTTGGGTAGATCATCATCCTTACCTAGTGGGCTTCCGACCTCATGGGGTACATT 60  
Db 1 ATGTTGGGTAGATCATCATCCTTACCTAGTGGGCTTCCGACCTCATGGGGTACATT 60  
Qy 61 CCGCTGCTGGGGCCCCCTAGGGGGTGTCCAGAGCCCTGGCGCATGGCGTGGGTT 120  
61 CCGCTGCTGGGGCCCCCTAGGGGGTGTCCAGAGCCCTGGCGCATGGCGTGGGTT 120  
Db 61 CCGCTGCTGGGGCCCCCTAGGGGGTGTCCAGAGCCCTGGCGCATGGCGTGGGTT 120  
Qy 121 CTGGAAGAGGGGCTGAACATGCAACAGGAAATTTGCTTCTCTATCTTC 180  
121 CTGGAAGAGGGGCTGAACATGCAACAGGAAATTTGCTTCTCTATCTTC 180  
Db 121 CTGGAAGAGGGGCTGAACATGCAACAGGAAATTTGCTTCTCTATCTTC 180  
Qy 181 CTTTGGCTTTACTGCTGCTGCAACATTCAGCTTCGGCTTATGAGGTGCGCAACGTG 240  
181 CTTTGGCTTTACTGCTGCTGCAACATTCAGCTTCGGCTTATGAGGTGCGCAACGTG 240  
Db 181 CTTTGGCTTTACTGCTGCTGCAACATTCAGCTTCGGCTTATGAGGTGCGCAACGTG 240  
Qy 241 TCCGGGATGTACCATGTCAAGACGACTGCTCCAACTCAAGCATTTGTATGAGGAGCG 300  
241 TCCGGGATGTACCATGTCAAGACGACTGCTCCAACTCAAGCATTTGTATGAGGAGCG 300  
Db 241 TCCGGGATGTACCATGTCAAGACGACTGCTCCAACTCAAGCATTTGTATGAGGAGCG 300  
Qy 301 GACATGATCATGACACACCCCGGGTGGCCCTGGTCCGGAGAACAACTTCCCGC 360  
301 GACATGATCATGACACACCCCGGGTGGCCCTGGTCCGGAGAACAACTTCCCGC 360  
Db 301 GACATGATCATGACACACCCCGGGTGGCCCTGGTCCGGAGAACAACTTCCCGC 360  
Qy 361 TCGTGGGTAGCCCTACCCCGAGCTCGAGTAGAAGCCAGCGTCCCACTAGAGACA 420  
361 TCGTGGGTAGCCCTACCCCGAGCTCGAGTAGAAGCCAGCGTCCCACTAGAGACA 420  
Db 361 TCGTGGGTAGCCCTACCCCGAGCTCGAGTAGAAGCCAGCGTCCCACTAGAGACA 420  
Qy 421 ATACGACGCCACGTGATTTGCTGTTGGGGGGGCTTTTCTGTTCCGCTATGACGTG 480  
421 ATACGACGCCACGTGATTTGCTGTTGGGGGGGCTTTTCTGTTCCGCTATGACGTG 480  
Db 421 ATACGACGCCACGTGATTTGCTGTTGGGGGGGCTTTTCTGTTCCGCTATGACGTG 480  
Qy 481 GGGGATCTGGGATCTGCTTCCGCTGCTCCAGCTTACCACTTCAGCTGCGCGG 540  
481 GGGGATCTGGGATCTGCTTCCGCTGCTCCAGCTTACCACTTCAGCTGCGCGG 540  
Db 481 GGGGATCTGGGATCTGCTTCCGCTGCTCCAGCTTACCACTTCAGCTGCGCGG 540  
Qy 541 CATGAGACGCTCAGAGACTGCAATTCATATCCCGGCACATTAACAGGTACCGT 600  
541 CATGAGACGCTCAGAGACTGCAATTCATATCCCGGCACATTAACAGGTACCGT 600  
Db 541 CATGAGACGCTCAGAGACTGCAATTCATATCCCGGCACATTAACAGGTACCGT 600  
Qy 601 ATGGCTGGGATATGATGATGAATGGTATATAG 633

Db 601 ATGGCTGGGATATGATGATGAATGGTATATAG 633

RESULT 4  
LOCUS A48675  
DEFINITION Sequence 13 from Patent WO9604385.  
ACCESSION A48675  
VERSION A48675.1 GI:2302388  
KEYWORDS  
SOURCE  
ORGANISM  
REFERENCE  
1 (bases 1 to 636)  
Maertens,G., Bosman,F., De M.G. and Buysse,M.  
TITLE PURIFIED HEPATITIS C VIRUS ENVELOPE PROTEINS FOR DIAGNOSTIC AND THERAPEUTIC USE  
JOURNAL Patent: WO 9604385-A 13 15-FEB-1996;  
INNOGENETICS NV (BE)  
COMMENT Other publication CA 2172273 960215  
Other publication AU 3382495 960304.  
FEATURES  
source  
1. 636  
Location/Qualifiers  
/organism="unidentified"  
/db\_xref="taxon:32644"  
1. .633  
CDS  
/note="unnamed protein product"  
/codon\_start=1  
/protein\_id="CAA03133.1"  
/db\_xref="GI:2302389"  
/translation="MLGRAIDITLCGFADLMGYIDLVGAPLGGAAALAHGVRLDGVNATGNLPGCSFSLFLALSLCTIPASAEVRNVSQMYHTNDSSNSIYEADIMHTPGCVPCVRENSSRCWVALPTPLAARNASVPTTIRRHVDLLVGAAPCSAMYGDIGSVFLVSQLFTISPRRHETVQDCNSIYPGHITGHRAMDMMMW"

mat\_peptide  
1. 630  
/product="unnamed"

BASE COUNT 111 a 197 c 175 g 153 t

ORIGIN

Query Match 96.8%; Score 612.8; DB 6; Length 636;  
Best Local Similarity 98.1%; Pred. No. 5.3e-139; Mismatches 12; Indels 0; Gaps 0;  
Matches 620; Conservative 0;

Db 1 ATGTTGGGTAGATCATCATCCTTACCTAGTGGGCTTCCGACCTCATGGGGTACATT 60  
1 ATGTTGGGTAGATCATCATCCTTACCTAGTGGGCTTCCGACCTCATGGGGTACATT 60  
Db 1 ATGTTGGGTAGATCATCATCCTTACCTAGTGGGCTTCCGACCTCATGGGGTACATT 60  
Qy 61 CCGCTGCTGGGGCCCCCTAGGGGGTGTCCAGAGCCCTGGCGCATGGCGTGGGTT 120  
61 CCGCTGCTGGGGCCCCCTAGGGGGTGTCCAGAGCCCTGGCGCATGGCGTGGGTT 120  
Db 61 CCGCTGCTGGGGCCCCCTAGGGGGTGTCCAGAGCCCTGGCGCATGGCGTGGGTT 120  
Qy 121 CTGGAAGAGGGGCTGAACATGCAACAGGAAATTTGCTTCTCTATCTTC 180  
121 CTGGAAGAGGGGCTGAACATGCAACAGGAAATTTGCTTCTCTATCTTC 180  
Db 121 CTGGAAGAGGGGCTGAACATGCAACAGGAAATTTGCTTCTCTATCTTC 180  
Qy 181 CTTTGGCTTTACTGCTGCTGCAACATTCAGCTTCGGCTTATGAGGTGCGCAACGTG 240  
181 CTTTGGCTTTACTGCTGCTGCAACATTCAGCTTCGGCTTATGAGGTGCGCAACGTG 240  
Db 181 CTTTGGCTTTACTGCTGCTGCAACATTCAGCTTCGGCTTATGAGGTGCGCAACGTG 240  
Qy 241 TCCGGGATGTACCATGTCAAGACGACTGCTCCAACTCAAGCATTTGTATGAGGAGCG 300  
241 TCCGGGATGTACCATGTCAAGACGACTGCTCCAACTCAAGCATTTGTATGAGGAGCG 300  
Db 241 TCCGGGATGTACCATGTCAAGACGACTGCTCCAACTCAAGCATTTGTATGAGGAGCG 300  
Qy 301 GACATGATCATGACACACCCCGGGTGGCCCTGGTCCGGAGAACAACTTCCCGC 360  
301 GACATGATCATGACACACCCCGGGTGGCCCTGGTCCGGAGAACAACTTCCCGC 360  
Db 301 GACATGATCATGACACACCCCGGGTGGCCCTGGTCCGGAGAACAACTTCCCGC 360  
Qy 361 TCGTGGGTAGCCCTACCCCGAGCTCGAGTAGAAGCCAGCGTCCCACTAGAGACA 420  
361 TCGTGGGTAGCCCTACCCCGAGCTCGAGTAGAAGCCAGCGTCCCACTAGAGACA 420  
Db 361 TCGTGGGTAGCCCTACCCCGAGCTCGAGTAGAAGCCAGCGTCCCACTAGAGACA 420  
Qy 421 ATACGACGCCACGTGATTTGCTGTTGGGGGGGCTTTTCTGTTCCGCTATGACGTG 480

```

Db 421 ATACACACCACAGTCGATTTGCTGTTGGGGGCGCTGCTTCTGTTCCGATATGACGTG 480
OY 481 GGGGATCTCTCGGAGTCTGTCTCTGTCCTCCAGCTGTTCACATCTCCGCTCGCCGG 540
Db 481 GGGGATCTCTCGGAGTCTGTCTCTGTCCTCCAGCTGTTCACATCTCCGCTCGCCGG 540
OY 541 CATGAGACGGTGCAGAGACTGCAATTGCTCAATCTATCCCGGCACATATACAGTCCCGT 600
Db 541 CATGAGACGGTGCAGAGACTGCAATTGCTCAATCTATCCCGGCACATATACAGTCCCGT 600
OY 601 ATGCGTTGGATATGATGATGAATGATTA 632
Db 601 ATGCGTTGGATATGATGATGAATGATTA 632

```

```

RESULT 5
ARI57329
LOCUS ARI57329 636 bp DNA Linear PAT 17-OCT-2001
DEFINITION Sequence 13 from patent US 6245503
ACCESSION ARI57329
VERSION ARI57329.1 GI:16218262
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS Maertens,G., Bosman,F., De Martynoff,G. and Buysse,M.-A.
TITLE Purified hepatitis C virus envelope proteins for diagnostic and therapeutic use
JOURNAL Patent: US 6245503-A 13 12-JUN-2001;
FEATURES
source location/Qualifiers
1..636
BASE COUNT 111 a 197 c 175 g 153 t
ORIGIN

```

```

Query Match 96.8%; Score 612.8; DB 6; Length 636;
Best Local Similarity 98.1%; Pred. No. 5.3e-139;
Matches 620; Conservative 0; Mismatches 12; Indels 0; Gaps 0;

```

```

OY 1 ATGTGGGTAAGAGTATGATACCTTACGTGCGCTTCGCCGACCTCATGGGGTACATT 60
Db 1 ATGCTGGGTAAGAGTATGATACCTTACGTGCGCTTCGCCGACCTCATGGGGTACATT 60
OY 61 CGGCTGTCGCGGCCCGCTAGGGGGTGTGTCAGAGCCCTGCGCATGGCGGGGT 120
Db 61 CGGCTGTCGCGGCCCGCTAGGGGGTGTGTCAGAGCCCTGCGCATGGCGGGGT 120
OY 121 CTGGAAGACGGCGTGAATATGCAACAGGGAATTTGCTGTTGCTTCTATCTTTC 180
Db 121 CTGGAAGACGGCGTGAATATGCAACAGGGAATTTGCTGTTGCTTCTATCTTTC 180
OY 181 CTCTGGCTTACTGCTGCTGTGACCAATTCAGCTTCGCTTATAGAGTGGCAACGTG 240
Db 181 CTCTGGCTTACTGCTGCTGTGACCAATTCAGCTTCGCTTATAGAGTGGCAACGTG 240
OY 241 TCCGGGATGTACATGTGTCAGAACAGTGTCTCAACTCAAGCATTTGTATAGAGAGCG 300
Db 241 TCCGGGATGTACATGTGTCAGAACAGTGTCTCAACTCAAGCATTTGTATAGAGAGCG 300
OY 301 GACATGATCATGACACACCCCGGTCGCTGCTGCTGCGGAGAACACTTTCGCCG 360
Db 301 GACATGATCATGACACACCCCGGTCGCTGCTGCTGCGGAGAACACTTTCGCCG 360
OY 361 TGGTGGGTAAGCGCTACCCCGACGCTGCGAGTAGAGAACGCCAGGCTCCCACTACAGACA 420
Db 361 TGGTGGGTAAGCGCTACCCCGACGCTGCGAGTAGAGAACGCCAGGCTCCCACTACAGACA 420
OY 421 ATACGACGCGACGTCGATTTGCTGTTGGGGGCGCTGCTTCTGTTGCTGCTATGTAACGTG 480
Db 421 ATACGACGCGACGTCGATTTGCTGTTGGGGGCGCTGCTTCTGTTGCTGCTATGTAACGTG 480

```

```

OY 481 GGGGATCTCTCGGAGTCTGTCTCTGTCCTCCAGCTGTTCACATCTCCGCTCGCCGG 540
Db 481 GGGGATCTCTCGGAGTCTGTCTCTGTCCTCCAGCTGTTCACATCTCCGCTCGCCGG 540
OY 541 CATGAGACGGTGCAGAGACTGCAATTGCTCAATCTATCCCGGCACATATACAGTCCCGT 600
Db 541 CATGAGACGGTGCAGAGACTGCAATTGCTCAATCTATCCCGGCACATATACAGTCCCGT 600
OY 601 ATGCGTTGGATATGATGATGAATGATTA 632
Db 601 ATGCGTTGGATATGATGATGAATGATTA 632

```

```

RESULT 6
AX452762
LOCUS AX452762 636 bp DNA Linear PAT 06-JUL-2002
DEFINITION Sequence 13 from Patent EP1211315.
ACCESSION AX452762
VERSION AX452762.1 GI:21712447
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS Maertens,G., Bosman,F., de Martynoff,G. and Buysse,M.A.
TITLE Recombinant vectors for producing hcv envelope proteins
JOURNAL Patent: EP 1211315-A 13 05-JUN-2002;
FEATURES
source location/Qualifiers
1..636
/organism="Hepatitis C virus"
/db_xref="taxon:11103"
1..633
/note="unnamed protein product"
/codon_start=1
/protein_10="CAD38071.1"
/db_xref="GI:21712448"
/translaction="MLGKADITLFCGFADLVGYIPLVGAFLGNAARALAHGVRVLEDM
VNVATGNLPGCSFSLFILAILSLCTIPASAEVENVAGMHTVDCNSLVEAEADY
IMHPGCVPCYRENNSSRCRWALPTLAARRASFPITTRRHVDLYGAALFCSAMTY
GDILGSAFLVLSQLEFTTISPRRRETVQDNCSTYPRGHTTGHRRAMDMNNWY"

```

```

CDS
1..633
/note="unnamed protein product"
/codon_start=1
/protein_10="CAD38071.1"
/db_xref="GI:21712448"
/translaction="MLGKADITLFCGFADLVGYIPLVGAFLGNAARALAHGVRVLEDM
VNVATGNLPGCSFSLFILAILSLCTIPASAEVENVAGMHTVDCNSLVEAEADY
IMHPGCVPCYRENNSSRCRWALPTLAARRASFPITTRRHVDLYGAALFCSAMTY
GDILGSAFLVLSQLEFTTISPRRRETVQDNCSTYPRGHTTGHRRAMDMNNWY"

```

```

mat_peptide 1..630
/product="unnamed"
BASE COUNT 111 a 197 c 175 g 153 t
ORIGIN
Query Match 96.8%; Score 612.8; DB 6; Length 636;
Best Local Similarity 98.1%; Pred. No. 5.3e-139;
Matches 620; Conservative 0; Mismatches 12; Indels 0; Gaps 0;

```

```

OY 1 ATGTGGGTAAGAGTATGATACCTTACGTGCGCTTCGCCGACCTCATGGGGTACATT 60
Db 1 ATGCTGGGTAAGAGTATGATACCTTACGTGCGCTTCGCCGACCTCATGGGGTACATT 60
OY 61 CGGCTGTCGCGGCCCGCTAGGGGGTGTGTCAGAGCCCTGCGCATGGCGGGGT 120
Db 61 CGGCTGTCGCGGCCCGCTAGGGGGTGTGTCAGAGCCCTGCGCATGGCGGGGT 120
OY 121 CTGGAAGACGGCGTGAATATGCAACAGGGAATTTGCTGTTGCTTCTATCTTTC 180
Db 121 CTGGAAGACGGCGTGAATATGCAACAGGGAATTTGCTGTTGCTTCTATCTTTC 180
OY 181 CTCTGGCTTACTGCTGCTGTGACCAATTCAGCTTCGCTTATAGAGTGGCAACGTG 240
Db 181 CTCTGGCTTACTGCTGCTGTGACCAATTCAGCTTCGCTTATAGAGTGGCAACGTG 240
OY 241 TCCGGGATGTACATGTGTCAGAACAGTGTCTCAACTCAAGCATTTGTATAGAGAGCG 300
Db 241 TCCGGGATGTACATGTGTCAGAACAGTGTCTCAACTCAAGCATTTGTATAGAGAGCG 300
OY 241 TCCGGGATGTACATGTGTCAGAACAGTGTCTCAACTCAAGCATTTGTATAGAGAGCG 300
Db 241 TCCGGGATGTACATGTGTCAGAACAGTGTCTCAACTCAAGCATTTGTATAGAGAGCG 300
OY 301 GACATGATCATGACACACCCCGGTCGCTGCTGCTGCGGAGAACACTTTCGCCG 360
Db 301 GACATGATCATGACACACCCCGGTCGCTGCTGCTGCGGAGAACACTTTCGCCG 360

```



QY 181 CTTCTGGCTTACTGCTCTGCTGACCATTCAGCTTCGCTTATGAGTGGCAACGTG 240  
DB 181 CTTCTGGCTTACTGCTCTGCTGACCATTCAGCTTCGCTTATGAGTGGCAACGTG 240  
QY 241 TCCGGGATGTACATGTGACGAGACAGTGTCTCAACTCAAGCATTTGTATGAGCAGCG 300  
DB 241 TCCGGGATGTACATGTGACGAGACAGTGTCTCAACTCAAGCATTTGTATGAGCAGCG 300  
QY 301 GACATGATCATGACACACCCCGGGTGGCTGCGCTTCGCGGAGAACAACTCTTCCCGC 360  
DB 301 GACATGATCATGACACACCCCGGGTGGCTGCGCTTCGCGGAGAACAACTCTTCCCGC 360  
QY 361 TCGTGGGTAGGCGCTCACCCCGCAGCTCGAGTAGAAGCCAGCGTCCCACTACAGACA 420  
DB 361 TCGTGGGTAGGCGCTCACCCCGCAGCTCGAGTAGAAGCCAGCGTCCCACTACAGACA 420  
QY 421 ATACAGCCGACGTCGATTTGCTGCTGTTGGGGGGGCTGCTTTCTGTTCCGCTATGACGTG 480  
DB 421 ATACAGCCGACGTCGATTTGCTGCTGTTGGGGGGGCTGCTTTCTGTTCCGCTATGACGTG 480  
QY 481 GGGGATCTCTGGGATCTGCTCTCTCTGCTCCAGCTGGTTCACCATCTCCGCTCGCCG 540  
DB 481 GGGGATCTCTGGGATCTGCTCTCTCTGCTCCAGCTGGTTCACCATCTCCGCTCGCCG 540  
QY 541 CATGAGAGGCTGACAGACTGCAATTGCTCAATCTATCCGGCCACATTAACAGTCAAGCT 600  
DB 541 CATGAGAGGCTGACAGACTGCAATTGCTCAATCTATCCGGCCACATTAACAGTCAAGCT 600  
QY 601 ATGGCTTGGGATATGATGATGAAGTGT 628  
DB 601 ATGGCTTGGGATATGATGATGAAGTGT 628

RESULT 9  
AX452754 795 bp DNA linear PAT 06-JUL-2002  
LOCUS AX452754  
DEFINITION Sequence 5 from Patent EP1211315.  
ACCESSION AX452754  
VERSION AX452754.1 GI:21712439  
KEYWORDS  
SOURCE  
ORGANISM  
Hepatitis C virus.  
Hepatitis C virus  
Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;  
Hepadnavirus.

REFERENCE  
1  
AUTHORS Maertens,G., Bosman,F., de Martynoff,G. and Buyse,M.A.  
TITLE Recombinant vectors for producing hcv envelope proteins  
JOURNML Patent: EP 1211315-A 5 05-JUN-2002;  
Innogenetics N.V. (BE)  
FEATURES  
source  
1..795  
/organism="Hepatitis C virus"  
/db\_xref="taxon:11103"  
1..792  
/note="unnamed protein product"  
/codon\_start=1  
/protein\_id="CAD38067.1"  
/db\_xref="GI:21712440"  
/translation="MLGKVIDTLICGFADLVGYIPLVGAPLGGAAALAHGVRVLEDC  
VNATGNLPGCSFSIFLIALSLTLPASAVEVRVSGMYHTNDSCNSIYEAADM  
IMHTPGCTPCYRENNSSRCRWALPTLAARASVPTTIRRHVLDLVAAPFCSMAY  
GDLGCVFLVSQLEFTISPRHEIVDDCSITFGHTTSHRAMDMKMMWSPTTALVVS  
QLRIPOAVVMVAGAHWGLAGLAVYSWVGNMAKVLIVLLFAP"  
mat\_peptide  
1..789  
/product="unnamed"  
BASE COUNT 130 a 240 c 231 g 194 t  
ORIGIN  
Query Match 96.2%; Score 608.8; DB 6; Length 795;  
Best Local Similarity 98.1%; Pred. No. 5,1e-138;  
Matches 616; Conservative 0; Mismatches 12; Indels 0; Gaps 0;

QY 1 ATGTTGGGTAAAGTCAATGATCACTTACGTCGGGCTTCGCCGACCTCATGGGGTACATT 60

DB 1 ATGTTGGGTAAAGTCAATGATCACTTACGTCGGGCTTCGCCGACCTCATGGGGTACATT 60  
QY 61 CCGCTGTCGGGGCCCCCTTAGGGGGTCTGCCAAGAGCCCTGGCCATAGGCGTCCGGGTT 120  
DB 61 CCGCTGTCGGGGCCCCCTTAGGGGGTCTGCCAAGAGCCCTGGCCATAGGCGTCCGGGTT 120  
QY 121 CTGAAGACGGCGGTGAACATGACAGAGGAAATTTGCTGCTGCTTCTATATCTTC 180  
DB 121 CTGAAGACGGCGGTGAACATGACAGAGGAAATTTGCTGCTGCTTCTATATCTTC 180  
QY 181 CTTCTGGCTTACTGCTCTGCTGACCATTCAGCTTCGCTTATGAGTGGCAACGTG 240  
DB 181 CTTCTGGCTTACTGCTCTGCTGACCATTCAGCTTCGCTTATGAGTGGCAACGTG 240  
QY 241 TCCGGGATGTACATGTGACGAGACAGTGTCTCAACTCAAGCATTTGTATGAGCAGCG 300  
DB 241 TCCGGGATGTACATGTGACGAGACAGTGTCTCAACTCAAGCATTTGTATGAGCAGCG 300  
QY 301 GACATGATCATGACACACCCCGGGTGGCTGCGCTTCGCGGAGAACAACTCTTCCCGC 360  
DB 301 GACATGATCATGACACACCCCGGGTGGCTGCGCTTCGCGGAGAACAACTCTTCCCGC 360  
QY 361 TCGTGGGTAGGCGCTCACCCCGCAGCTCGAGTAGAAGCCAGCGTCCCACTACAGACA 420  
DB 361 TCGTGGGTAGGCGCTCACCCCGCAGCTCGAGTAGAAGCCAGCGTCCCACTACAGACA 420  
QY 421 ATACAGCCGACGTCGATTTGCTGCTGTTGGGGGGGCTGCTTTCTGTTCCGCTATGACGTG 480  
DB 421 ATACAGCCGACGTCGATTTGCTGCTGTTGGGGGGGCTGCTTTCTGTTCCGCTATGACGTG 480  
QY 481 GGGGATCTCTGGGATCTGCTCTCTCTGCTCCAGCTGGTTCACCATCTCCGCTCGCCG 540  
DB 481 GGGGATCTCTGGGATCTGCTCTCTCTGCTCCAGCTGGTTCACCATCTCCGCTCGCCG 540  
QY 541 CATGAGAGGCTGACAGACTGCAATTGCTCAATCTATCCGGCCACATTAACAGTCAAGCT 600  
DB 541 CATGAGAGGCTGACAGACTGCAATTGCTCAATCTATCCGGCCACATTAACAGTCAAGCT 600  
QY 601 ATGGCTTGGGATATGATGATGAAGTGT 628  
DB 601 ATGGCTTGGGATATGATGATGAAGTGT 628

RESULT 10  
A48709 2082 bp DNA linear PAT 07-MAR-1997  
LOCUS A48709  
DEFINITION Sequence 47 from Patent WO9604385.  
ACCESSION A48709  
VERSION A48709.1 GI:2302422  
KEYWORDS  
SOURCE  
ORGANISM  
unidentified.  
unidentified  
unclassified.  
1 (bases 1 to 2082)  
REFERENCE  
1  
AUTHORS Maertens,G., Bosman,F., De,M.G. and Buyse,M.  
TITLE PURIFIED HEPATITIS C VIRUS ENVELOPE PROTEINS FOR DIAGNOSTIC AND  
THERAPEUTIC USE  
JOURNAL Patent: WO 9604385-A 47 15-FEB-1996;  
INNOMERGENICS NV (BE)  
COMMENT Other publication CA 2172273 960215  
Other publication AU 3382495 960304.  
FEATURES  
source  
1..2082  
/organism="unidentified"  
/db\_xref="taxon:32644"  
1..2079  
/note="unnamed protein product; Protein sequence is in  
conflict with the conceptual translation"  
/codon\_start=1  
/protein\_id="CAA03146.1"  
/db\_xref="GI:2302423"  
/translation="MLGKVIDTLICGFADLVGYIPLVGAPLGGAAALAHGVRVLEDC

VNYATGNLPCCSFIFLLALISLTVPPASATVENVNCGMTHVTNDCSNSSIYEADMD  
 IMHTGCVPCVRENNSRCWALTPPLAARNASVPTT11RRHVDLLVGAACASAMV  
 GDLCSVEFVQLFTISPRHETVODNCSTPGHTRHMAAMMNSPTTALVVS  
 QLLRIPOAVVDMVAGAHGVLAGLAYSVAGVAVLVYMLFAGVDITRVSAGAA  
 SDRGLVSLFSPGSOKIOLVNTNGSMHINPRLNCDNSIQGFPAALPYKRRKSSG  
 CPERLASCRSIDKFAOGKPLATTPRPNSSDDQPYCWHAPRRCGLVPSOVQVPCF  
 TPSPVVGTDTREVPYTNMGANDSDVILLNTRPPRGWFCCTMMNGFPRKTCGGP  
 PCNIGAGNNMTLTCPTDCFRKHEATVARGSPWLTPECMVHYPRILMHYPCVNTF  
 IEFKMYVGVGVEHREACNMTRGERCDLEDBRSELSPLLSTEMOILLPCSFETLP  
 ALSTGLIHONHIVDOYLYGVSAVSVLITWEYVLEILLADARICACIMMLLI  
 NOAEALENLVYLNMAAAYAGANGSLTFLVFPICAMV IKGRLYPGAAYFYGWPLILL  
 LLALPPRAVA"  
 1..2076  
 /product="unnamed"

mat\_peptide  
 BASE COUNT 366 a 634 c 600 g 482 t  
 ORIGIN

Query Match 95.7%; Score 605.8; DB 6; Length 2082;  
 Best Local Similarity 98.1%; Pred. No. 2.9e-137;

Matches 613; Conservative 0; Mismatches 12; Indels 0; Gaps 0;

OY 4 TTGGGTAAAGTCAATGATACCTTACGTGGGGTTCGCCGACCTCATGGGATCATTCG 63  
 DB 4 TTGGGTAAAGTCAATGATACCTTACGTGGGGTTCGCCGACCTCATGGGATCATTCG 63  
 OY 64 CTGCTGGGGCCCCCCTAGGGGGTCTGCCAGAGCCCTGGGCATGGCGTCCGGTTCG 123  
 DB 64 CTGCTGGGGCCCCCCTAGGGGGTCTGCCAGAGCCCTGGGCATGGCGTCCGGTTCG 123  
 OY 124 GAAGACGGCGTCAATGATGCAACAGGGAATTTGCTGCTTCTCTATCTTCCTC 183  
 DB 124 GAAGACGGCGTCAATGATGCAACAGGGAATTTGCTGCTTCTCTATCTTCCTC 183  
 OY 184 TTGGCTTACTGTCGTGTCGACCATTCAGCTTCGGCTATAGAGGCGCAACGTCGC 243  
 DB 184 TTGGCTTACTGTCGTGTCGACCATTCAGCTTCGGCTATAGAGGCGCAACGTCGC 243  
 OY 244 GGGATGATGATGCAACAGGAGCTGCTCCAACTCAAGCATTTGTATAGAGGAGCGGAC 303  
 DB 244 GGGATGATGATGCAACAGGAGCTGCTCCAACTCAAGCATTTGTATAGAGGAGCGGAC 303  
 OY 304 ATGATGATGCAACAGGAGCTGCTCCAACTCAAGCATTTGTATAGAGGAGCGGAC 303  
 DB 304 ATGATGATGCAACAGGAGCTGCTCCAACTCAAGCATTTGTATAGAGGAGCGGAC 303  
 OY 364 TTGGGTAGCGGTCAACCCGAGCTGAGGAGCCAGCGTCCCACTAGACAAATA 423  
 DB 364 TTGGGTAGCGGTCAACCCGAGCTGAGGAGCCAGCGTCCCACTAGACAAATA 423  
 OY 424 CGACGCCAGGTGATTTGCTGCTGGGGGCGCTTTCCTTCCCTATGTAAGTGGG 483  
 DB 424 CGACGCCAGGTGATTTGCTGCTGGGGGCGCTTTCCTTCCCTATGTAAGTGGG 483  
 OY 484 GATCTCTGGGATCTGCTCTGCTGCTCCAGCTGTTCACCATCTTCGCTCCCGGAT 543  
 DB 484 GATCTCTGGGATCTGCTCTGCTGCTCCAGCTGTTCACCATCTTCGCTCCCGGAT 543  
 OY 544 GAGACGGTGAGGAGTGCATATGCTCAATCTATCCGGGCACATAACAGGTCAACGATG 603  
 DB 544 GAGACGGTGAGGAGTGCATATGCTCAATCTATCCGGGCACATAACAGGTCAACGATG 603  
 OY 604 GCTTGGGATATGATGATGAAGTGGT 628  
 DB 604 GCTTGGGATATGATGATGAAGTGGT 628

RESULT 11  
 ARI57350  
 LOCUS ARI57350 2082 bp DNA linear PAT 17-OCT-2001  
 DEFINITION Sequence 47 from patent US 6245503.  
 ACCESSION ARI57350  
 VERSION ARI57350.1 GI:16218284  
 KEYWORDS

SOURCE Unknown.  
 ORGANISM Unknown.  
 UNCLASSIFIED.  
 REFERENCE 1 (bases 1 to 2082)  
 AUTHORS Maertens, G., Bosman, F., De Martynoff, G. and Buyse, M.-A.  
 TITLE Purified hepatitis C virus envelope proteins for diagnostic and therapeutic use  
 JOURNAL Patent: US 6245503-A 47 12-JUN-2001;  
 FEATURES location/Qualifiers  
 source 1..2082  
 BASE COUNT 366 a 634 c 600 g 482 t  
 ORIGIN

Query Match 95.7%; Score 605.8; DB 6; Length 2082;  
 Best Local Similarity 98.1%; Pred. No. 2.9e-137;

Matches 613; Conservative 0; Mismatches 12; Indels 0; Gaps 0;

OY 4 TTGGGTAAAGTCAATGATACCTTACGTGGGGTTCGCCGACCTCATGGGATCATTCG 63  
 DB 4 TTGGGTAAAGTCAATGATACCTTACGTGGGGTTCGCCGACCTCATGGGATCATTCG 63  
 OY 64 CTGCTGGGGCCCCCCTAGGGGGTCTGCCAGAGCCCTGGGCATGGCGTCCGGTTCG 123  
 DB 64 CTGCTGGGGCCCCCCTAGGGGGTCTGCCAGAGCCCTGGGCATGGCGTCCGGTTCG 123  
 OY 124 GAAGACGGCGTCAATGATGCAACAGGGAATTTGCTGCTTCTCTATCTTCCTC 183  
 DB 124 GAAGACGGCGTCAATGATGCAACAGGGAATTTGCTGCTTCTCTATCTTCCTC 183  
 OY 184 TTGGCTTACTGTCGTGTCGACCATTCAGCTTCGGCTATAGAGGCGCAACGTCGC 243  
 DB 184 TTGGCTTACTGTCGTGTCGACCATTCAGCTTCGGCTATAGAGGCGCAACGTCGC 243  
 OY 244 GGGATGATGATGCAACAGGAGCTGCTCCAACTCAAGCATTTGTATAGAGGAGCGGAC 303  
 DB 244 GGGATGATGATGCAACAGGAGCTGCTCCAACTCAAGCATTTGTATAGAGGAGCGGAC 303  
 OY 304 ATGATGATGCAACAGGAGCTGCTCCAACTCAAGCATTTGTATAGAGGAGCGGAC 303  
 DB 304 ATGATGATGCAACAGGAGCTGCTCCAACTCAAGCATTTGTATAGAGGAGCGGAC 303  
 OY 364 TTGGGTAGCGGTCAACCCGAGCTGAGGAGCCAGCGTCCCACTAGACAAATA 423  
 DB 364 TTGGGTAGCGGTCAACCCGAGCTGAGGAGCCAGCGTCCCACTAGACAAATA 423  
 OY 424 CGACGCCAGGTGATTTGCTGCTGGGGGCGCTTTCCTTCCCTATGTAAGTGGG 483  
 DB 424 CGACGCCAGGTGATTTGCTGCTGGGGGCGCTTTCCTTCCCTATGTAAGTGGG 483  
 OY 484 GATCTCTGGGATCTGCTCTGCTGCTCCAGCTGTTCACCATCTTCGCTCCCGGAT 543  
 DB 484 GATCTCTGGGATCTGCTCTGCTGCTCCAGCTGTTCACCATCTTCGCTCCCGGAT 543  
 OY 544 GAGACGGTGAGGAGTGCATATGCTCAATCTATCCGGGCACATAACAGGTCAACGATG 603  
 DB 544 GAGACGGTGAGGAGTGCATATGCTCAATCTATCCGGGCACATAACAGGTCAACGATG 603  
 OY 604 GCTTGGGATATGATGATGAAGTGGT 628  
 DB 604 GCTTGGGATATGATGATGAAGTGGT 628

RESULT 12  
 AX452796  
 LOCUS AX452796 2082 bp DNA linear PAT 06-JUL-2002  
 DEFINITION Sequence 47 from Patent EP1211315.  
 ACCESSION AX452796  
 VERSION AX452796.1 GI:21712481  
 KEYWORDS  
 SOURCE Hepatitis C virus.  
 ORGANISM Hepatitis C virus  
 Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;

## Hepacivirus.

REFERENCE 1  
Maertens,G., Bosman,F., de Martynoff,G. and Buysse,M.A.  
AUTHORS  
Recombinant vectors for producing hcv envelope proteins  
TITLE  
Patent: EP 1211315-A 47 05-JUN-2002;  
JOURNAL  
Innogenetics N.V. (BE)

## FEATURES

## source

1..2082

Location/Qualifiers

## CDS

/organism="Hepatitis C virus"

/db\_xref="taxon:11103"

&lt;1..2079

/note="unnamed protein product"

/codon\_start=1

/protein\_id="CAD38082.1"

/db\_xref="GI:21712482"

/translation="NLGKVIDITLTCGFADLVGYIPLVGAPLGAARALAHGVLEEDG  
VNYATGNLPGCSFSEIFLLALISCLTVPASAEVNSQMYHTNDCSSSYEADIMH  
IPGCVPCVRENSSRCWALPTLARAASVPTTIRRHVDLLVGAAPCSAMVY  
GDLGSAVPLVSQLEFTISPRHETVODNCSTYPRGHITGHRAMDMNMSPTTALVYS  
QLRIPOAVDMVAGAHMGLAVLITSVGNMAKVLVNLVLFAGVDGHTRVSGAA  
SDTRGLVSLFSPGSAOKIQLVNTNGSMHINTALNCDSLOTGEFALFYHKFNSSG  
CPEPLASCRSIDRFAOGMGLTYTEBNSDQRPYCMHYAPPGIIVPASOYGVYCF  
TSPVYVGTDRFGVPTNMGANDSVLLINNTBPRGNMFGCTMMNGTGTGCGP  
PCNIGGAGNNLTGCTPCDCEKREHPEATYARCGSGPMTPRCMVHYPYRLMHPCTVNF  
IFKRYMYGVGEHREPRACANTRGERCDLERDSELSPLSTTEMQILPCSTTLLP  
ALSTGLIHLHONIVDOYLYGVSAVSVLTKEVYLLFLLELADARICACLMMLLI  
AQAALALENLVNLVNLAAVAGAHGTLSELVFECAMVYIKGRLLVPGAAYVAYGVPLLL  
LLALPPRAYA"

## mat\_peptide

1..2076

/product="unnamed"

BASE COUNT 366 a 634 c 600 g 482 t

## ORIGIN

## Query Match

Best Local Similarity 98.1%; Score 605.8; DB 6; Length 2082;  
Matches 613; Conservative 0; Mismatches 12; Indels 0; Gaps 0;

0Y 4 TTGGTAAGGTATGATACCTTACGTCGGCTTCCGACCTATGGGGTACATTCG 63  
|||||  
Db 4 TTGGTAAGGTATGATACCTTACGTCGGCTTCCGACCTATGGGGTACATTCG 63  
|||||  
0Y 64 CTGCTGCGGGCCCCCTAGGGGGGTGCTGCAGAGCCCTGGCCATGGGGTTCG 123  
|||||  
Db 64 CTGCTGCGGGCCCCCTAGGGGGGTGCTGCAGAGCCCTGGCCATGGGGTTCG 123  
|||||  
0Y 124 GAAGACGGCGTGAACATATGCAACAGGGAATTTGCCGTGTTGCTTCTCTATCTTCTC 183  
|||||  
Db 124 GAAGACGGCGTGAACATATGCAACAGGGAATTTGCCGTGTTGCTTCTCTATCTTCTC 183  
|||||  
0Y 184 TTGGCTTTACTGCTGCTGTCTGACCAATTCACGCTTCCGCTTATGAGTGGCAACGTGTC 243  
|||||  
Db 184 TTGGCTTTACTGCTGCTGTCTGACCAATTCACGCTTCCGCTTATGAGTGGCAACGTGTC 243  
|||||  
0Y 244 GGGATGTCACATGTCAGCAAGACGTCGTCACACTCAAGATTTGTATAGGACAGCGAC 303  
|||||  
Db 244 GGGATGTCACATGTCAGCAAGACGTCGTCACACTCAAGATTTGTATAGGACAGCGAC 303  
|||||  
0Y 304 ATGATCATGACACACCCCGGGGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 363  
|||||  
Db 304 ATGATCATGACACACCCCGGGGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 363  
|||||  
0Y 364 TGGGTAGGCGTCAACCCCAAGCTGCAAGTGAAGAACGCCAGCGTCCCACTACGACAAATA 423  
|||||  
Db 364 TGGGTAGGCGTCAACCCCAAGCTGCAAGTGAAGAACGCCAGCGTCCCACTACGACAAATA 423  
|||||  
0Y 424 CGAGCCCAAGTCGATTTGCTGTTGGGGGGGTGCTGTTGCTGTTGCTGTTGCTGTTGCTGTTG 483  
|||||  
Db 424 CGAGCCCAAGTCGATTTGCTGTTGGGGGGGTGCTGTTGCTGTTGCTGTTGCTGTTGCTGTTG 483  
|||||  
0Y 484 GATCTCTGGGATCTGCT 543  
|||||  
Db 484 GATCTCTGGGATCTGCT 543  
|||||

0Y 544 GAGACGGTCAGAGTGCATATTCATCTATCCCGCCACATACAGCTCACCTATG 603  
|||||  
Db 544 GAGACGGTCAGAGTGCATATTCATCTATCCCGCCACATACAGCTCACCTATG 603  
|||||  
0Y 604 GCTTGATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 628  
|||||  
Db 604 GCTTGATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 628  
|||||

## RESULT 13

## A48711

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

FEATURES

source

1..2433

/organism="unidentified"

/db\_xref="taxon:32644"

1..2430

/note="unnamed protein product"

/codon\_start=1

/protein\_id="CA03147.1"

/db\_xref="GI:2302425"

/translation="MSTNPKFQRTKTRNTNRPODYKFPGGQIVGVYLLIPRGPPL  
GVATRTKTSEROPRGRROPTKARREPRGAWAOPGTPWPLYGEGMGWAGWLLSPG  
SRPWSGPTDPRRRSRNLGKVIDLTLCGFADLVGYIPLVGAPLGAARALAHGVLEED  
GVNYATGNLPGCSFSEIFLLALISCLTVPASAEVNSQMYHTNDCSSSYEADIMH  
IPGCVPCVRENSSRCWALPTLARAASVPTTIRRHVDLLVGAAPCSAMVY  
GDLGSAVPLVSQLEFTISPRHETVODNCSTYPRGHITGHRAMDMNMSPTTALVYS  
QLRIPOAVDMVAGAHMGLAVLITSVGNMAKVLVNLVLFAGVDGHTRVSGAA  
ASDTRGLVSLFSPGSAOKIQLVNTNGSMHINTALNCDSLOTGEFALFYHKFNSSG  
CPEPLASCRSIDRFAOGMGLTYTEBNSDQRPYCMHYAPPGIIVPASOYGVYCF  
TSPVYVGTDRFGVPTNMGANDSVLLINNTBPRGNMFGCTMMNGTGTGCGP  
PCNIGGAGNNLTGCTPCDCEKREHPEATYARCGSGPMTPRCMVHYPYRLMHPCTVNF  
IFKRYMYGVGEHREPRACANTRGERCDLERDSELSPLSTTEMQILPCSTTLLP  
ALSTGLIHLHONIVDOYLYGVSAVSVLTKEVYLLFLLELADARICACLMMLLI  
AQAALALENLVNLVNLAAVAGAHGTLSELVFECAMVYIKGRLLVPGAAYVAYGVPLLL  
LLALPPRAYA"

## mat\_peptide

1..2427

/product="unnamed"

BASE COUNT 434 a 745 c 714 g 540 t

## ORIGIN

## Query Match

Best Local Similarity 95.7%; Score 605.8; DB 6; Length 2433;  
Matches 613; Conservative 0; Mismatches 12; Indels 0; Gaps 0;

0Y 4 TTGGTAAGGTATGATACCTTACGTCGGCTTCCGACCTATGGGGTACATTCG 63  
|||||  
Db 355 TTGGTAAGGTATGATACCTTACGTCGGCTTCCGACCTATGGGGTACATTCG 63  
|||||  
0Y 64 CTGCTGCGGGCCCCCTAGGGGGGTGCTGCAGAGCCCTGGCCATGGGGTTCG 123  
|||||  
Db 415 CTGCTGCGGGCCCCCTAGGGGGGTGCTGCAGAGCCCTGGCCATGGGGTTCG 123  
|||||  
0Y 124 GAAGACGGCGTGAACATATGCAACAGGGAATTTGCCGTGTTGCTTCTCTATCTTCTC 183  
|||||  
Db 475 GAAGACGGCGTGAACATATGCAACAGGGAATTTGCCGTGTTGCTTCTCTATCTTCTC 183  
|||||  
0Y 184 TTGGCTTTACTGCTGCTGTCTGACCAATTCACGCTTCCGCTTATGAGTGGCAACGTGTC 243  
|||||

```

|||||
Db 535 TTGGCTTGGCTGCTCTGCTGACCCGTTCCAGCTTCCGTTATGAAAGTGGCAAGCTGTCC 594
OY 244 GGGATGTACCATGTGACGAAGCACTGCTCCAACTCAAGCAATTGATGAGAGCGAGAC 303
Db 595 GGGATGTACCATGTGACGAAGCACTGCTCCAACTCAAGCAATTGATGAGAGCGAGAC 654
OY 304 ATGATCATGTGACACACCCCGGGGTGGTCCCTGCTTCCGGAGAAACAATCTTCCCGTGC 363
Db 655 ATGATCATGTGACACACCCCGGGGTGGTCCCTGCTTCCGGAGAAACAATCTTCCCGTGC 714
OY 364 TTGGTACCGCTCAACCCCGAGCTGCGAGCTAGGAAGCCAGCCGCTCCACTACGACATA 423
Db 715 TTGGTACCGCTCAACCCCGAGCTGCGAGCTAGGAAGCCAGCCGCTCCACTACGACATA 774
OY 424 CGAGCCACGCTCGAATTTGCTGCTGGGGGGGCTGCTTCTGTTCCGCTATGATGAGTGGG 483
Db 775 CGAGCCACGCTCGAATTTGCTGCTGGGGGGGCTGCTTCTGTTCCGCTATGATGAGTGGG 834
OY 484 GATCTCTGGGAGATCTGCTCTCTGCTCTCCAGCTGTTCAACCATCTGCGCTCGCCGCGCAT 543
Db 835 GACCTCTGGGAGATCTGCTCTCTGCTCTCCAGCTGTTCAACCATCTGCGCTCGCCGCGCAT 894
OY 544 GAGACGGTGCAGAGCTCAATTTGCTCAATCTATCCCGCCGACATAACAGCTCACCGTATG 603
Db 895 GAGACGGTGCAGAGCTCAATTTGCTCAATCTATCCCGCCGACATAACAGCTCACCGTATG 954
OY 604 GCTTGGATATGATGATGATGACTGCT 628
Db 955 GCTTGGATATGATGATGATGACTGCT 979

```

```

RESULT 14
LOCUS AR157351 2433 bp DNA linear PAT 17-OCT-2001
DEFINITION Sequence 49 from patent US 6245503.
ACCESSION AR157351
VERSION AR157351.1 GI:16218285
KEYWORDS
SOURCE
ORGANISM
REFERENCE
1 (bases 1 to 2433)
Maertens,G., Bosman,F., De Martynoff,G. and Buyse,M.-A.
Purified hepatitis C virus envelope proteins for diagnostic and
therapeutic use
Patent: US 6245503-A 49 12-JUN-2001;
FEATURES
source location/Qualifiers
1..2433
BASE COUNT 434 a 745 c 714 g 540 t
ORIGIN

```

Query Match 95.7%; Score 605.8; DB 6; Length 2433;  
 Best Local Similarity 98.1%; Pred. No. 2,9e-137;  
 Matches 613; Conservative 0; Mismatches 12; Indels 0; Gaps 0;

```

OY 4 TTGGGTAGAGTCAATCATGATACCTTACGTTGCGGCTTCCGACGCTCATGGGTACATCCG 63
Db 355 TTGGGTAGAGTCAATCATGATACCTTACGTTGCGGCTTCCGACGCTCATGGGTACATCCG 414
OY 64 CTCGTGCGCGCCCGCCCTAGGGGGTGTCTGCCAGAGCCCTGGCGCATGGCGTCCGGGTTCTG 123
Db 415 CTCGTGCGCGCCCGCCCTAGGGGGTGTCTGCCAGAGCCCTGGCGCATGGCGTCCGGGTTCTG 474
OY 124 GAGAGCGCGCTGAACATTTGCAACAGGAATTTGCTGCTGCTTCTCTATCTTCTCTC 183
Db 475 GAGAGCGCGCTGAACATTTGCAACAGGAATTTGCTGCTGCTTCTCTATCTTCTCTC 534
OY 184 TTGGCTTACTGCTGCTGCTGACCATTTCCAGCTTATGAGTGGCCAAAGTGTCC 243
Db 535 TTGGCTTACTGCTGCTGCTGACCATTTCCAGCTTATGAGTGGCCAAAGTGTTC 594

```

```

OY 244 GGGATGTACCATGTGACGAAGCACTGCTCCAACTCAAGCAATTGATGAGAGCGAGAC 303
Db 595 GGGATGTACCATGTGACGAAGCACTGCTCCAACTCAAGCAATTGATGAGAGCGAGAC 654
OY 304 ATGATCATGTGACACACCCCGGGGTGGTCCCTGCTTCCGGAGAAACAATCTTCCCGTGC 363
Db 655 ATGATCATGTGACACACCCCGGGGTGGTCCCTGCTTCCGGAGAAACAATCTTCCCGTGC 714
OY 364 TTGGTACCGCTCAACCCCGAGCTGCGAGCTAGGAAGCCAGCCGCTCCACTACGACATA 423
Db 715 TTGGTACCGCTCAACCCCGAGCTGCGAGCTAGGAAGCCAGCCGCTCCACTACGACATA 774
OY 424 CGAGCCACGCTCGAATTTGCTGCTGGGGGGGCTGCTTCTGTTCCGCTATGATGAGTGGG 483
Db 775 CGAGCCACGCTCGAATTTGCTGCTGGGGGGGCTGCTTCTGTTCCGCTATGATGAGTGGG 834
OY 484 GATCTCTGGGAGATCTGCTCTCTGCTCTCCAGCTGTTCAACCATCTGCGCTCGCCGCGCAT 543
Db 835 GACCTCTGGGAGATCTGCTCTCTGCTCTCCAGCTGTTCAACCATCTGCGCTCGCCGCGCAT 894
OY 544 GAGACGGTGCAGAGCTCAATTTGCTCAATCTATCCCGCCGACATAACAGCTCACCGTATG 603
Db 895 GAGACGGTGCAGAGCTCAATTTGCTCAATCTATCCCGCCGACATAACAGCTCACCGTATG 954
OY 604 GCTTGGATATGATGATGATGACTGCT 628
Db 955 GCTTGGATATGATGATGATGACTGCT 979

```

```

RESULT 15
LOCUS AX452798 2433 bp DNA linear PAT 06-JUL-2002
DEFINITION Sequence 49 from Patent EP1211315.
ACCESSION AX452798
VERSION AX452798.1 GI:21712483
KEYWORDS
SOURCE
ORGANISM
REFERENCE
1
Maertens,G., Bosman,F., de Martynoff,G. and Buyse,M.A.
Recombinant vectors for producing hcv envelope proteins
Patent: EP 1211315-A 49 05-JUN-2002;
FEATURES
source location/Qualifiers
1..2433
CDS
1..2430
/organism="Hepatitis C virus"
/db_xref="taxon:11103"
/note="unnamed protein product"
/codon_start=1
/protein_id="CAD38083.1"
/db_xref="GI:21712483"
/translating="MSTNPKPQKTRKRNTRRRPQDKFFGGQIVGGVYLLPRGRRL
GVRATRKTSEROSPRGRQPIPKARRPBRRAAOQPYMPPLGDNBMGAGWLLSPRG
SRPMGPTDPRRRSRMLGVIDLTFCGADLVGYIPLVGAIPGGAARALAHVRLIED
GVNATAGNLPGCSFEFLALISCLTVPASAVEVRNVCAMVHTVNDSSSTVTAAD
MIMHPGCVPCVRENNSRCWALPTTLAARNAVPTTIRRHVDLLVGAAPFCAMV
VDLGCSEVLEVSOLFISPRRHEVVDACSITPGHITGHRAMDMNMSPTTALVY
SOLLRPOAVVMVAGAHGVLGLAYSMVGNMAKVLVVMILFAGVDHRTVSGAAV
ASDTRGLVLSFPGSAOKIOLVNTNGSMHINFTALCNDISLOTGFPAALFYRHKFNS
GCPERLASCRSIDKFNQMGWPLTYTEPESDORPCMHYAPRPGIVAPSOVCGPVYC
FPPSPVVGTTDPRFGVPTYNMGANDSDVLLNNTPRRGNMGCTMMNGTGFTKCGG
PPCNIGGAGNNMTLTCPTDCEFRKHPATVYRCSGRPLTPRCAVHPIRYLRMHPTCVNF
TTRKRMVYGVGVRPEACNMTRRGERCDLEDRDSSELSPLLSTTEWQIILPDSFTTL
PALSTGLHLHONIVDOYLVGVGSAVSVLVFKWEVILLFLFLADARICADLMMMLL
IAOERALENLVVLNAAVAAGAHGTLFLVFCAMV INGRVLPGAAYAFYGVMPDLL
LILALPPRAYA"
1..2427
mat_peptide
/peptide="unnamed"

```

```

BASE COUNT 434 a 745 c 714 g 540 t
ORIGIN

```

Query Match	95.7%;	Score 605.8;	DB 6;	Length 2433;
Best Local Similarity	98.1%;	Pred. No. 2.9e-137;		
Matches 613; Conservative	0;	Mismatches 12;	Indels 0;	Gaps 0;

QY	4	TTGGTAAAGTCATCGATGATACCCCTTACGTGGGGCTTGCCGCAACCTGATGGGGTACATTCGG	63
Db	355	TTGGGTAAAGTCATCGATGATACCCCTTACATGGGGCTTGCCGCAACCTGAGGGGTACATTCGG	41.44
QY	64	CTCGTCGGCGCCCCCTTAAGGGGGTGTCTGCCAGAGACCCCTGGCGCATGGAGCTCCGGTCTG	123
Db	415	CTCGTCGGCGCCCCCTTAAGGGGGGGTGTGCCAGAGGCCCTTGCGCATGGAGCTCCGGGTCTG	47.4
QY	124	GAAACAGGCGCTGAACCTATGCAACAGGGAAATTTCCTGTGGTGTCTTTCTCTATCTTCTC	183
Db	475	GAGACAGGCGCGTAACCTATGCAACAGGGAAATTTCGCCGTGCTCTTCTCTATCTTCTC	53.44
QY	184	TTGGCTTACTGCTCGTGTACCATTTCCAGCTTCGCGCTTATGAGGTGGCAACGTGTC	24.3
Db	535	TTGGCTTGTGCTGTCTGTCTCTACCGCTTCCAGCTTCGCGCTTATGAAGTGGCAAGGTGTC	59.4
QY	244	GGGATGTACCATGTTCAGGAACGACTGCTCCAACTTGTGTATGAGCGAGCGAC	30.3
Db	595	GGGATGTACCATGTTCAGGAACGAGTGTCTCCAACTGATGTATGAGCGAGCGAC	65.4
QY	304	ATGATCATGACACACCCCGGGTGGGTGCCCTTGCGCTTTGGGGAAGAACACTCTTCCCGTGC	36.5
Db	655	ATGATCATGACACACCCCGGGGTGGTGCCTTCGCTTGGGGAAGAACACTCTTCCCGTGC	71.4
QY	364	TGGGTAGCGCTCACCCCCACGCTGCGAGTATAGAAACGACGACGTCACCTATACGACATA	42.3
Db	715	TGGGTAGCGCTCACCCCCACGCTGCGAGCTATAGAAAGCGACGTCACCTACGACACATA	77.4
QY	424	CGACGCCACGTCGATTTTGCTGTTGGGGCGGCTGCTTGTCTCCGCTATATGACGTGGG	48.3
Db	775	CGAGGCCACGTCGATTTTGCTGTTGGGGCGGCTGCTTGTCTCCGCTATATGACGTGGG	83.4
QY	484	GATCTATCGGATATGATCTTCTGTGTCCAGCTGTTCACATCTGACGCTCGCGGCAT	54.3
Db	835	GACCTCTGCGATCTGTCTGTGTCTCCAGCTGTTCACATCTGACGCTCGCGGCAT	89.4
QY	544	GAGACGGTGCAGGACATGCAATTTGCTCATCTATCCCGGCAATATACGATCAACCTATG	60.3
Db	895	GAGACGGTGCAGGACATGCAATTTGCTCATCTATCCCGGCAATATACGATCAACCTATG	95.4
QY	604	GCTTGGGATATGATGATGACTGGT	628
Db	955	GCTTGGGATATGATGATGACTGGT	979

Search completed: February 18, 2003, 06:39:08  
Job time : 2149 secs

GenCore version 5.1.3  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - protein search, using frame\_plus\_n2p model

Run on: February 12, 2003, 08:55:00 ; Search time 39 Seconds  
(without alignments)  
4325.520 Million cell updates/sec

Title: US-09-899-303a-7  
Sequence: 1 ATGTTGGTGAAGCATCATCA.....TGATGATGACGTGTAATAG 633

Scoring table:  
BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 1816940

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters:  
-MODE=frame-n2p.model -DEV=x1p  
-Q=cn2.1/USPTO.spool/US0989303/rtunal.12022003.085448.28253/app.query.fasta\_1.775  
-DB=A.Geneseq.101002 -QFMT=fastan -SUFFIX=n2p.rag -MINMATCH=0.1 -LOOPEXT=0  
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blomsun62 -TRANS=humand0.coi  
-LIST=45 -DOCALLIGN=200 -THR\_SCORE=pct -THR\_MAX=100 -THR\_MIN=0 -ALIGN=15  
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US0989303.eccn.1.1.30.0rtunal.12022003.085448.28253 -NCPU=6 -ICPU=3  
-NO\_XIPXY -NO\_MMAP -LARGEQUERY -NEG\_SCORES=0 -WAIT -LONGLOG -DEV\_TIMEOUT=120  
-NARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7  
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :  
1: A.Geneseq.101002.\*  
2: /SID52/gcgdata/geneseq/genesqp-emb1/AA1980.DAT.\*  
3: /SID52/gcgdata/geneseq/genesqp-emb1/AA1981.DAT.\*  
4: /SID52/gcgdata/geneseq/genesqp-emb1/AA1983.DAT.\*  
5: /SID52/gcgdata/geneseq/genesqp-emb1/AA1984.DAT.\*  
6: /SID52/gcgdata/geneseq/genesqp-emb1/AA1985.DAT.\*  
7: /SID52/gcgdata/geneseq/genesqp-emb1/AA1986.DAT.\*  
8: /SID52/gcgdata/geneseq/genesqp-emb1/AA1987.DAT.\*  
9: /SID52/gcgdata/geneseq/genesqp-emb1/AA1988.DAT.\*  
10: /SID52/gcgdata/geneseq/genesqp-emb1/AA1989.DAT.\*  
11: /SID52/gcgdata/geneseq/genesqp-emb1/AA1990.DAT.\*  
12: /SID52/gcgdata/geneseq/genesqp-emb1/AA1991.DAT.\*  
13: /SID52/gcgdata/geneseq/genesqp-emb1/AA1992.DAT.\*  
14: /SID52/gcgdata/geneseq/genesqp-emb1/AA1993.DAT.\*  
15: /SID52/gcgdata/geneseq/genesqp-emb1/AA1994.DAT.\*  
16: /SID52/gcgdata/geneseq/genesqp-emb1/AA1995.DAT.\*  
17: /SID52/gcgdata/geneseq/genesqp-emb1/AA1997.DAT.\*  
18: /SID52/gcgdata/geneseq/genesqp-emb1/AA1997.DAT.\*  
19: /SID52/gcgdata/geneseq/genesqp-emb1/AA1998.DAT.\*  
20: /SID52/gcgdata/geneseq/genesqp-emb1/AA1998.DAT.\*  
21: /SID52/gcgdata/geneseq/genesqp-emb1/AA2000.DAT.\*  
22: /SID52/gcgdata/geneseq/genesqp-emb1/AA2001.DAT.\*  
23: /SID52/gcgdata/geneseq/genesqp-emb1/AA2002.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1095	94.8	310	13	AAR20720
2	1094	94.7	729	19	AA447264
3	1094	94.7	733	19	AA447264
4	1087	94.1	513	13	AAR24087
5	1084	93.9	340	15	AA49537
6	1081	93.6	729	19	AA447265
7	1081	93.6	1031	15	AA45067
8	1081	93.6	1031	16	AA45067
9	1081	93.6	3010	22	AA98022
10	1081	93.6	3010	22	AA98022
11	1081	93.6	3010	22	AA98022
12	1081	93.6	3011	20	AA98021
13	1075	93.1	3010	22	AA98021
14	1073	92.9	416	13	AA25854
15	1073	92.9	445	13	AA25854
16	1073	92.9	603	12	AA12600
17	1073	92.9	980	15	AA53921
18	1073	92.9	1010	17	AA88311
19	1073	92.9	3010	16	AA82694
20	1073	92.9	3010	16	AA82694
21	1073	92.9	3010	16	AA82694
22	1070	92.6	470	14	AA34479
23	1070	92.6	470	14	AA34480
24	1069	92.6	416	13	AA24526
25	1069	92.6	465	13	AA21685
26	1067	92.4	560	14	AA32589
27	1067	92.4	3011	13	AA34468
28	1066	92.3	326	19	AA68461
29	1066	92.3	423	19	AA68461
30	1066	92.3	845	19	AA68461
31	1065	92.2	224	19	AA68461
32	1065	92.2	555	15	AA68461
33	1065	92.2	555	16	AA72800
34	1065	92.2	1051	15	AA54066
35	1065	92.2	1051	16	AA54066
36	1063	92.0	321	13	AA25115
37	1063	92.0	470	14	AA34469
38	1063	92.0	470	14	AA34476
39	1063	92.0	470	14	AA34477
40	1063	92.0	470	14	AA34478
41	1063	92.0	3010	14	AA30616
42	1060	91.8	3010	13	AA20091
43	1060	91.8	3010	13	AA20111
44	1060	91.8	3010	20	AA706423
45	1059	91.7	470	14	AA34467

## ALIGNMENTS

## RESULT 1

ID AAR20720 standard; Protein; 310 AA.

AC AAR20720:

DT 06-MAY-1992 (first entry)

DE C10-E12 NANBH-specific antigen polypeptide.

XX Non-A non-B hepatitis virus; recombinant; detection.

OS Non-A non-B hepatitis virus.

XX EP468657-A.

PD 29-JAN-1992.

PF 08-JUL-1991; 91BP-0306158.







Db 159 GluaspelValaInTyralaThrglyAsnLeuProglyCysSerPheSerIlePheLeu 178  
 QY 184 TTGGCTTTACTGCTCTGTCACCATTCACGCTTCCGCTTATGAGTGCGCAACGTGTCC 243  
 Db 179 LeuValleuLeuSerCysLeuThrIleProAlaSerAlaTyrgluValaArgAsnValSer 198  
 QY 244 GGGATGACCATGTCAGAGAACGACTGCTCCAACTCAAGCATTTGTTATGAGGACGGGAC 303  
 Db 199 GlyIleTyrlHisValThrAsnAspCysSerAsnSerIleValTyrgluAlaAlaAsp 218  
 QY 304 ATGATCATGCACACCCCGGGTGGCTGCCGCTTCCGGAGAACACTGTTCCGGTGC 363  
 Db 219 ValIleMetHisAlaProglyCysValProCysValaArggluAsnAsnSerSerArgCys 238  
 QY 364 TGGGTACCGCTCAACCCCGCTGCGAGCTAGAGAACCCGCGCTCCCACTACGACATA 423  
 Db 239 TrpValaIleuThrProThrIleuAlaAlaArgAsnAlaSerValProThrThrIleu 258  
 QY 424 CGAGCGCAGCTGATTTGGCTGCTGGGGCGGCTTCTTCTTCCGCTATGATGAGTGGG 483  
 Db 259 ArgArgHisValAspLeuValaGlyThrAlaAlaPheCysSerAlaMetTyrlValGly 278  
 QY 484 GATCTGTCGCGATCTGCTCTGCTGCTCCAGCTGTTCAACATCTGCGCGCGGCGAT 543  
 Db 279 AspLeuCyseglySerValaIlePheLeuIleSerGlnLeuPheThrPheSerProAlaArgHis 298  
 QY 544 GAGACGGTGCAGAGCTGCAATTTGCTCAATCTATCCCGCCACATACAGGTCACCGTATG 603  
 Db 299 GluThrValaGlnAspCysAsnCysSerIleTyrlProGlyHisValSerGlyHisArgMet 318  
 QY 604 GCTTGGATATGATGATGAACCTGG 627  
 Db 319 AlaTrpAspMetMetMetAsnTrp 326

RESULT 6  
 AAM47265  
 ID AAM47265 standard; Protein: 729 AA.  
 AC AAM47265;  
 DT 15-MAR-2002 (first entry)  
 DE Cuticle protein 1 and 2 secreting hepatitis C virus related protein #2.  
 XX Cuticle protein 1; cuticle protein 2; hepatitis C virus.  
 KW Unidentified.  
 OS  
 XX KR97065713-A.  
 PN 13-OCT-1997.  
 PD 19-MAR-1996; 96KR-0007404.  
 PF 19-MAR-1996; 96KR-0007404.  
 XX 19-MAR-1996; 96KR-0007404.  
 PR (GLDS ) LG CHEM LTD.  
 PA Choo SH, Lee IH, Ryo WS;  
 PI MPI: 1998-492654/42.  
 DR N-PSDB; ABA03492.  
 XX N-PSDB; ABA03492.  
 PT Cuticle protein 1 and 2 secreting hepatitis C virus (Japanese)  
 PT Noabstract -  
 XX  
 PS Disclosure: Page 4-5; 7pp: Korean.  
 CC The present invention relates to cuticle protein 1 and 2 secreting  
 CC hepatitis C virus. The present sequence is a protein provided in the  
 CC exemplification of the invention.  
 XX  
 SO Sequence 729 AA;

Alignment Scores:  
 Pred. No.: 3,33e-104 Length: 729  
 Score: 1081.00 Matches: 199  
 Percent Similarity: 99.04% Conservative: 7  
 Best Local Similarity: 95.67% Mismatches: 2  
 Query Match: 93.59% Indels: 0  
 DB: 19 Gaps: 0

US-09-899-303a-7 (1-633) x AAM47265 (1-729)

QY 4 TTGGGTAAAGTCAATCATACCTTACGTCGGCGCTTCCGCGACCTCATGGGATACATTCCG 63  
 Db 119 LeuGlyIleValIleAspThrIleuThrCysGlyPheAlaAspLeuMetGlyTyrlPro 138  
 QY 64 CTCGTCGGCCCGCCCGTAGGGGGTGTGTCGACAGCCCTGGCGGATGGCGTCGGGTTCTG 123  
 Db 139 LeuValaGlyAlaProLeuGlyGlyAlaAlaArgAlaLeuAlaHisGlyValaIleValLeu 158  
 QY 124 GAAGACGGCGTGAACATAATGCAACAGGGAATTGCTGCTGCTTCTTCTATCTTCTCTC 183  
 Db 159 GluaspelValaInTyralaThrglyAsnLeuProglyCysSerPheSerIlePheLeu 178  
 QY 184 TTGGCTTTACTGCTCTGTCACCATTCACGCTTCCGCTTATGAGTGCGCAACGTGTCC 243  
 Db 179 LeuValleuLeuSerCysLeuThrIleProAlaSerAlaTyrgluValaArgAsnValSer 198  
 QY 244 GGGATGACCATGTCAGAGAACGACTGCTCCAACTCAAGCATTTGTTATGAGGACGGGAC 303  
 Db 199 GlyIleTyrlHisValThrAsnAspCysSerAsnSerIleValTyrgluAlaAlaAsp 218  
 QY 304 ATGATCATGCACACCCCGGGTGGCTGCCGCTTCCGGAGAACACTGTTCCGGTGC 363  
 Db 219 ValIleMetHisAlaProglyCysValProCysValaArggluAsnAsnSerSerArgCys 238  
 QY 364 TGGGTACCGCTCAACCCCGCTGCGAGCTAGAGAACCCGCGCTCCCACTACGACATA 423  
 Db 239 TrpValaIleuThrProThrIleuAlaAlaArgAsnAlaSerValProThrThrIle 258  
 QY 424 CGAGCGCAGCTGATTTGGCTGCTGGGGCGGCTTCTTCTTCCGCTATGATGAGTGGG 483  
 Db 259 ArgArgHisValAspLeuValaGlyThrAlaAlaPheCysSerAlaMetTyrlValGly 278  
 QY 484 GATCTGTCGCGATCTGCTCTGCTGCTCCAGCTGTTCAACATCTGCGCGCGGCGAT 543  
 Db 279 AspLeuCyseglySerValaIlePheLeuIleSerGlnLeuPheThrPheSerProAlaArgHis 298  
 QY 544 GAGACGGTGCAGAGCTGCAATTTGCTCAATCTATCCCGCCACATACAGGTCACCGTATG 603  
 Db 299 GluThrValaGlnAspCysAsnCysSerIleTyrlProGlyHisValSerGlyHisArgMet 318  
 QY 604 GCTTGGATATGATGATGAACCTGG 627  
 Db 319 AlaTrpAspMetMetMetAsnTrp 326

RESULT 7  
 AAR54067  
 ID AAR54067 standard; Protein: 1031 AA.  
 AC AAR54067;  
 DT 14-FEB-1995 (first entry)  
 DE Non-A, non-B hepatitis virus gene #6 product.  
 XX  
 DE Non-A, non-B hepatitis virus gene #6 product.  
 XX  
 KW Non-A, non-B hepatitis virus; NANBHV; hepatitis C virus; HCV;  
 KW core; ENV; NS1; NS2; NS3; antigen; detection.  
 XX  
 OS Hepatitis C virus.  
 XX  
 FH Key Location/Qualifiers  
 FT region 1..191  
 FT /label= core

FT	Region	192..383	/label= ENV
FT	Region	384..810	/label= NS1
FT	Region	811..1031	/label= NS2-NS3
FT			/note= "NS3 N-terminal"
XX			
PN	JF0614870-A.		
XX			
PD	24-MAY-1994.		
XX			
PF	12-MAR-1992;	92JP-0088140.	
XX			
PR	12-MAR-1992;	92JP-0088140.	
XX			
PA	(SANW ) SANWA KAGAKU KENKYUSHO CO.		
PA	(TOFU ) TONEN CORP.		
PA	(TOKR-) ZH TOKYO RINSHO IGAKU SOGO KENKYUSHO.		
XX			
DR	WPI: 1994-205028/25.		
XX	N-PSDB: AA064069.		
XX			
PT	DNA coding a Non-A, non-B hepatitis virus antigen - useful for		
PT	detecting HCV within serum		
XX			
PS	Claim 6-10; Page 15-19; 22pp; Japanese.		
XX			
CC	Hepatitis C virus #4 and #6 genes were isolated (AA064068-69).		
CC	Both genes contain the core, ENV, NS1, NS2 and NS3 regions.		
CC	A core region fragment is given in AA064067.		
XX			
SQ	Sequence 1031 AA:		
	Alignment Scores:		
	Pred. No.:	3,76e-104	Length: 1031
	Score:	1081.00	Matches: 199
	Percent Similarity:	98.08%	Conservative: 5
	Best Local Similarity:	95.67%	Mismatches: 4
	Query Match:	93.59%	Indels: 0
	DB:	15	Gaps: 0
	US-09-899-303A-7 (1-633) x AAR54067 (1-1031)		
OY	4 TTGGGTAAAGTGTCATGCATACCCCTTACGTGGCGGCTTGGCGGACCTCANGGGGTACATTCCG	63	
Db	119 LeuGIyLysValIleAspPThrLeuThrcysglYpheaIAspLeuMetGIyTyrIlePro	138	
OY	64 CTCGCGGGGGCCCCCGTATGGGGGGGTGCGCCAGACCCCGCGGCATGGCGTCCGGGTTCTG	123	
Db	139 LeuValGIyAlaProLeuGIyValAlaIArgAlaLeuAlaHisGIyValAlaIArgValVal	158	
OY	124 GAAAGACGGCGTAACATATGACAAGAGGATTTGGCTGGTCTTCTCTATCTTCTCTC	183	
Db	159 GluAspGIyValAsnTyrAlaIThrGIyAsnLeuProGIySsrPheSerIlePheLeu	178	
OY	184 TTGGCTTACTGTGCTGTCTGTACCATTCACGTTCCGCTTATGAGGTGCGCAAGTGTCC	243	
Db	179 LeuAlaIleuLeuSerCysLeuThrIleProAlaSerAlaTyrGIuValAlaIArgAsnValSer	198	
OY	244 GGGAGTATCCATGTGCAGCAAGCATGCTCCCACTCAAGCATTTGTGTATGAGCGACGGAC	303	
Db	199 GIyIleTyrHisValThrAsnAspCysSerAsnSerIleValTyrGIuAlaAlaAsp	218	
OY	304 ATGATCATGACACACCCCGGGGTGGTGGCGCTTGGCGGGAACAACATCTTCCCGCTGC	363	
Db	219 MetIleMetHisThrProGIyCysValProCysValAlaIArgGIuGIyAsnSerSerAlgCys	238	
OY	364 TGGGTAGCGCTACACCCCGACGCTCGACAGTATGGAAGCGCAAGCGTCCCGCAGTACGACATA	423	
Db	239 TrpValAlaIleuThrProThrLeuAlaIAlaIArgAsnAlaSerValProThrTrhAlaIle	258	
OY	424 CGACGCCACGTCGATTTGCTGTGGGGCGGCTCTTTCGTTCCGCTATATGTAAGTGGG	483	

D	b	.	259	ArgAgnHnSVaIaSPLeuLleValGlyAlaLaHaLaHeSySeSeraLaMeTyryVaIGLy	278
OY			484	GATCTGTCCGGATGTGTCCTTCCGTGCCTCCAGCTGTTCACCAATCTCGCTGCCCGGAT	543
D	b		279	AspleucyScgISerValPheLeuVaISerClnDeuBerhpRhesSerProAdgaRhnIs	298
OY			544	GAGACGGGCGAGCATGCAATTGGTCATCATTCACCSCGCCAACATAACAGTAACCCATATG	603
D	b		299	GIUThrlIEglnaSpCySaNcYSseTrlEtyrProSlYhISvaISergLYhlaRGmet	318
OY			604	GCTTGGAATGATGATGAATGACTGC	627
D	b		319	AlatrpaspMetMetMetasnTp	326
RESULT 8 AA98362					
ID			AA98362 standard; Protein; 1031 AA.		
XX			AA98362;		
AC					
XX			22-AUG-1996 (first entry)		
DT					
XX			'UTR/CORE/ENV/NS1/NS2/NS3 from HCV (#6).		
DE					
XX			Hepatitis C virus; HCV; antigen; detection; antibody.		
KM					
XX			Hepatitis C virus.		
OS					
Key			Location/Qualifiers		
FH			Peptide	1..191	
FT			Peptide	/label= core peptide	
FT			Peptide	192..363	
FT			Peptide	/label= ENV1	
FT			Peptide	364..810	
FT			Peptide	/label= NS1/NS2	
FT			Peptide	811..1031	
ET			Peptide	/label= NS2 and NS3	
XX			JPO713291-A.		
PX					
PD			23-MAY -1995.		
PF			18-JUN-1993; 93JP-0147944.		
PR			18-JUN-1993; 93JP-0147944.		
PA			(TOFU ) TONEN CORP.		
XX					
DR			WPI; 1995-220780/29.		
N-PSDB; AAT30387.					
Recombinant polypeptide comprising partial NS1 region of hepatitis non-A non-B viral antigen - used in a method for detecting antibodies against hepatitis non-A non-B virus.					
Disclosure; Page 13-15; 15pp; Japanese.					
The sequences given in AA98361-62 represent the 5'/UTR/CORE/ENV/NS1/NS2/NS3 protein region derived from hepatitis C virus (HCV) isolates #4 and #6 respectively. The proteins partic. contain amino acids 364-493 of the HCV NS1 antigen. These protein fragments may be used in the detection of antibodies against HCV.					
Sequence 1031 AA:					
Alignment Scores:					
Pred. NO.:			3.76e-104		1031
Score:			1081.00		Matches: 199
Percent Similarity:			98.08%		Conservative: 5
Best Local Similarity:			93.59%		Mismatches: 4
Query Match:			16		Indels: 0
aps:			0		aps: 0

US-09-899-303a-7 (1-633) x AAM98362 (1-1031)

QY 4 TTGGGTAAGTCATCATGATACCTTACGCTGCGGCTTCGCCGACCTCATGGGGTACATTCG 63  
 DB 119 LeucllyysValIleasphrhlleuthrhcysglyphealaspheumetgtyrillepro 138  
 QY 64 CTCGTCGCGCCGCCCTAGGCGGCTGCCAGAGCCCTGGCGCATGGCGGCTTCG 123  
 DB 139 LeuValIleAlaProleuGlyGlyValIleAlaArgAlaLeuAlaHISGlyValArgValVal 158  
 QY 124 GAAGACGGCGGTGACATTCGCAACAGGGAATTTGGCTGCTTCTTCTATCTTCTC 183  
 DB 159 GluaspIlyValAsnYrAlaThGlyAsnLeuProGlyCysSerPheSerIlePheleu 178  
 QY 184 TTGGCTTACTGCTGCTGACCATTCAGACTCCGCTTATGAGTGGCGACGCTGCC 243  
 DB 179 LeuAlaLeuLeuSerCysleuthrIleProAlaSerAlaTyrgIuValArgAsnValSer 198  
 QY 244 GCGATGATACATGTCACGAACGACTGCTCCAACTCAAGCATTTGTATGAGGACGCGAC 303  
 DB 199 GlyIleTyrlsValIleThraAsnAspCysSerAsnSerSerIleValTyrgIuAlaAlaAsp 218  
 QY 304 ATGATCATGACACACCCCGGGTGGCTGCTGCTGCTTGGGAGAACACTCTTCCGCTGC 363  
 DB 219 MetIleMetIleThrProGlyCysValProCysValArgGluGlyAsnSerSerArgCys 238  
 QY 364 TTGGTACGCTGACCCCGACGCTGCGAGCTAGAACGCGACGCTGCCACTAGACACATA 423  
 DB 239 TrpValAlaLeuThrProThrIleuAlaAlaArgAsnAlaSerValProThrThAlaIle 258  
 QY 424 CGACGCCACGTCGATTTGCTGCTGGGCGGCTGCTTCTGCTCCGTATGACGTGGG 483  
 DB 259 ArgArgIleValAspLeuLeuValGlyAlaAlaAlaPheCysSerAlaMetTyValGly 278  
 QY 484 GATCTGCGGAGATGCTCTCTCTGCTGCTCCAGCTGTCACATCTGGCTCCGCGGCAT 543  
 DB 279 AspLeuGlySerValPheLeuValSerClnLeuPheThrPheSerProArgArgHis 298  
 QY 544 GAGACGGTGCAGACTGCAATTTGCTCAATCATCCGCGCACATPACAGGTCACCGTATG 603  
 DB 299 GluThrIleGlnAspCysAsnGlySerIleTyrlProGlyHisValSerGlyHisArgMet 318  
 QY 604 GCTTGGATATGATGATGAACCTGG 627  
 DB 319 AlaTrpAspMetMetMetAsnTrp 326

RESULT 9  
 AAM98022 standard; protein; 3010 AA.  
 ID AAM98022:  
 AC AAM98022:  
 XX 21-JUN-1999 (first entry)  
 DT 21-JUN-1999 (first entry)  
 DE Infectious hepatitis C virus genotype 1b strain HC-J4 protein.  
 XX HCV; infectious clone; infection; diagnosis; therapy; vaccine;  
 KM screening; assay; antiviral; virucide.  
 XX Hepatitis C virus.  
 OS Hepatitis C virus.  
 XX MO9904008-A2.  
 XX 28-JAN-1999.  
 XX 16-JUL-1998; 98MO-US14688.  
 XX 27-JAN-1998; 98US-0014416.  
 XX 18-JUL-1997; 97US-0053062.  
 XX (USSH ) US DEPT HEALTH & HUMAN SERVICES.

PI Buhk J, Emerson SU, Purcell RH, Yanagi M;  
 XX WPI: 1999-132252/11.  
 DR N-PSDB; AAX24843.  
 XX  
 PT New isolated hepatitis C virus nucleic acids - used to develop  
 PT products for the diagnosis, prevention and treatment of HCV  
 PT infections and for developing screening assays  
 PS Claim 2; Fig 14G-H; 126pp; English.

CC This protein is encoded by the infectious hepatitis C virus (HCV)  
 CC genotype 1b strain HC-J4 genome (see AAX24833). HC-J4 was obtained  
 CC from acute phase plasma of a chimpanzee infected with serum  
 CC containing HC-J4/91. The infectious nucleic acid sequence can be  
 CC used to produce chimeric genomes (see AAX24833) consisting of the  
 CC open reading frames of infectious nucleic acid sequences of other  
 CC genotypes (including genotypes 1-6) and subtypes (such as 1b, 2a,  
 CC 2b, 2c, 3a, 4a-f, 5a and 6a) of HCV. The invention also relates to  
 CC the introduction of mutations or deletions into infectious nucleic  
 CC acid sequences in order to produce an attenuated HCV virus suitable  
 CC for vaccine development. Infectious nucleic acid sequences can also  
 CC be used to produce attenuated virus via passage in vitro or in vivo  
 CC of the viruses produced by transfection of a host cell with the  
 CC infectious nucleic acid sequence. Vaccines comprising one or more  
 CC polypeptides made from the infectious nucleic acid sequence are  
 CC used to immunise mammals, especially humans, against hepatitis C.  
 CC The nucleic acid sequences can also be used to induce protective  
 CC immunity against the virus. The nucleic acid sequences or their  
 CC encoded proteases (e.g. NS3 protease) can additionally be used to  
 CC develop screening assays to identify antiviral agents for HCV.

SQ Sequence 3010 AA:

Alignment Scores:  
 Pred. No.: 5,46e-104 Length: 3010  
 Score: 1081.00 Matches: 199  
 Percent Similarity: 98.56% Conservative: 6  
 Best Local Similarity: 95.67% Mismatches: 3  
 Query Match: 93.59% Indels: 0  
 DB: 20 Gaps: 0

US-09-899-303a-7 (1-633) x AAM98022 (1-3010)

QY 4 TTGGGTAAGTCATCATGATACCTTACGCTGCGGCTTCGCCGACCTCATGGGGTACATTCG 63  
 DB 119 LeucllyysValIleasphrhlleuthrhcysglyphealaspheumetgtyrillepro 138  
 QY 64 CTCGTCGCGCCGCCCTAGGCGGCTGCCAGAGCCCTGGCGCATGGCGGCTTCG 123  
 DB 139 LeuValIleAlaProleuGlyGlyValIleAlaArgAlaLeuAlaHISGlyValArgValVal 158  
 QY 124 GAAGACGGCGGTGACATTCGCAACAGGGAATTTGGCTGCTTCTTCTATCTTCTC 183  
 DB 159 GluaspIlyValAsnYrAlaThGlyAsnLeuProGlyCysSerPheSerIlePheleu 178  
 QY 184 TTGGCTTACTGCTGCTGACCATTCAGACTCCGCTTATGAGTGGCGACGCTGCC 243  
 DB 179 LeuAlaLeuLeuSerCysleuthrIleProAlaSerAlaTyrgIuValArgAsnValSer 198  
 QY 244 GCGATGATACATGTCACGAACGACTGCTCCAACTCAAGCATTTGTATGAGGACGCGAC 303  
 DB 199 GlyIleTyrlsValIleThraAsnAspCysSerAsnSerSerIleValTyrgIuAlaAlaAsp 218  
 QY 304 ATGATCATGACACACCCCGGGTGGCTGCTGCTGCTTGGGAGAACACTCTTCCGCTGC 363  
 DB 219 MetIleMetIleThrProGlyCysValProCysValArgGluGlyAsnSerSerArgCys 238  
 QY 364 TTGGTACGCTGACCCCGACGCTGCGAGCTAGAACGCGACGCTGCCACTAGACACATA 423  
 DB 239 TrpValAlaLeuThrProThrIleuAlaAlaArgAsnAlaSerValProThrThAlaIle 258  
 QY 424 CGACGCCACGTCGATTTGCTGCTGGGCGGCTGCTTCTGCTCCGTATGACGTGGG 483

```

Db 259 ARGAGNHSVALAspleuenuValGlyThrAlaAlaPheCysSerAlaMetYValGly 278
QY 484 GATCTCTGGGAGATCTGTTCTCTGCTCCAGCTGTCCACCATTCGCGCCGCGCAT 543
Db 279 AspleuCYsGlySerIlePheHeuValSerGlnLeuPheThrPheSerProAlaGlnHS 298
QY 544 GAGACGGTGCAGAGCTGCAATTGCTCAATCTATCCCGCCACATAACAGGTACCCGTATG 603
Db 299 GluthrValGlnAspCYsAsnCYsSerIleTyrProGlnHisValSerGlnHisArgMet 318
QY 604 GCTTGGGATATGATGATGAACCTGG 627
Db 319 AlATrPspMetMetMetAsnTrp 326

RESULT 10
AAB31170
ID AAB31170 standard; Protein; 3010 AA.
XX
AC AAB31170;
XX
DT 02-APR-2001 (first entry)
XX
DE Amino acid sequence of a hepatitis C virus (HCV) clone genotype 1b.
XX
KM Chimeric virus; bovine viral diarrhoea virus; BVDV; hepatitis C virus;
KM HCV; vaccine; viral inhibitor; antiviral.
XX
OS Hepatitis C virus.
XX
PN WO200075352-A2.
XX
PD 14-DEC-2000.
XX
PF 02-JUN-2000; 2000WO-US15527.
XX
PR 04-JUN-1999; 99US-0137817.
XX
PA (USSH ) US DEPT HEALTH & HUMAN SERVICES.
XX
PI Nam J, Bukh J, Emerson SU, Purcell RH;
XX
DR N-PSDB; AAC86939.
XX
PT WPI: 2001-071081/08.
XX
PT New nucleic acid comprising a chimeric bovine viral diarrhoea virus
PT genome in which the (non-)structural region has been replaced by
PT hepatitis C virus (HCV) genome useful for treating or preventing HCV
PT signs and symptoms -
XX
PS Disclosure: Fig 4G-H; 97p; English.
XX
XX The specification describes a nucleic acid comprising a chimeric virus
XX genome, specifically bovine viral diarrhoea virus (BVDV) genome in which
XX the (non-)structural region has been replaced by the (non-)structural
XX region of a hepatitis C virus (HCV) genome. The nucleic acids comprising
XX the chimeric virus and the chimeric virus are useful for identifying
XX cell lines capable of supporting the replication of these chimeric
XX viruses, in screening for neutralizing antibodies to HCV of different
XX genotypes, in the production of HCV-BVDV virions, for the development
XX of inactivated or attenuated vaccines to prevent HCV-BVDV in a mammal,
XX in studying the molecular properties of HCV indirectly in vitro, and in
XX identifying inhibitors of viral enzyme activity which would be useful
XX as antiviral agents. Formulations or compositions comprising the
XX chimeric virions may be used to treat or prevent the signs and symptoms
XX of HCV. The present sequence is encoded by a HCV clone, which is used
XX to construct chimeric nucleic acids of the invention.
XX
SQ Sequence 3010 AA;

```

Alignment Scores: 5.46e-104 Length: 3010  
 Pred. No.: 1081.00 Matches: 199  
 Score:

```

Percent Similarity: 98.56% Conservative: 6
Best Local Similarity: 95.67% Mismatches: 3
Query Match: 93.59% Indels: 0
DB: 22 Gaps: 0

us-09-899-303a-7 (1-633) x AAB31170 (1-3010)
QY 4 TTGGGTAAAGTCATGCAATGACCTTACAGTGGGCTTGGCCGACCTGAGGGGTACATTCGG 63
Db 119 LeuGlyLysValIleAspThrIleuThrIleuThrIleuThrIleuThrIleuThrIleu 138
QY 64 CTTCGTGGGCGCCCGCTAGGGGGGTGCTGCCAGAGCCCTGGCCGACGGCGGTCTG 123
Db 139 LeuValGlyAlaProIleuGlyAlaAlaArgAlaLeuAlaIleuAlaArgValIleu 158
QY 124 GAAGACGGCGTGAATATGACAGAGGATTTGCCGCTTGGCTTCTTCTTCTTCTTCTTCC 183
Db 159 GluAspGlyValAsnTyrAlaThrGlyAsnLeuProGlyCysSerPheSerIlePheLeu 178
QY 184 TTGGCTTTACTGTCTCTGCTGACCATTCAGCTTCCGCTTATGAGTGGCAACGTCTCC 243
Db 179 LeuAlaLeuLeuSerCysLeuThrIleProAlaSerAlaTyrGluValAlaGlnValSer 198
QY 244 GCGATGACCATGTCACGACGACAGCTGCTCAACTACAGCATTTGTATGAGCGAGCGAC 303
Db 199 GlyIleTyrHisValThrAsnAspCYsSerAsnSerSerIleValTyrGluAlaAlaAsp 218
QY 304 ATGATCATGACACACCCCGGGTGGCGCTGCTGGGGAACAACATCTCCGCTCC 363
Db 219 ValIleMetHisThrProIleCysValProCysValGlnGlyAsnSerSerArgCys 238
QY 364 TGGGTAGCGCTACCCCAAGCTCGAGCTAGGAAGCCAGCGCTCCCACTAGCAATA 423
Db 239 TrpValAlaLeuThrProThrIleuAlaAlaArgAlaAlaSerValProThrThrThrIle 258
QY 424 CGACGCCACGTGATTTGCTGCTGGGGCGGCTCTTGTGTCCGATATGACGTGGG 483
Db 259 ARGARGHSVALAspleuenuValGlyThrAlaAlaPheCysSerAlaMetYValGly 278
QY 484 GATCTCTGGGAGATCTGTTCTCTGCTCCAGCTGTCCACCATTCGCGCTCCGCGCAT 543
Db 279 AspleuCYsGlySerIlePheHeuValSerGlnLeuPheThrPheSerProAlaGlnHS 298
QY 544 GAGACGGTGCAGAGCTGCAATTGCTCAATCTATCCCGCCACATAACAGGTACCCGTATG 603
Db 299 GluthrValGlnAspCYsAsnCYsSerIleTyrProGlnHisValSerGlnHisArgMet 318
QY 604 GCTTGGGATATGATGATGAACCTGG 627
Db 319 AlATrPspMetMetMetAsnTrp 326

RESULT 11
AAE20477
ID AAE20477 standard; Protein; 3010 AA.
XX
AC AAE20477;
XX
DT 01-JUL-2002 (first entry)
XX
DE HCV-S1 full-length polypeptide.
XX
KM Nucleic acid construct; expression cassette; non-coding region; NCR;
KM untranslated region; UTR; anti-viral drug; drug resistance;
KM HCV-S1; Hepatitis C virus.
XX
OS Hepatitis C virus.
XX
PN WO200208447-A2.
XX
PD 31-JAN-2002.
XX
PF 20-JUL-2001; 2001WO-IL00669.
XX

```



CC develop screening assays to identify antiviral agents for HCV.

XX Sequence 3011 AA:

Alignment Scores:

Pred. No.:	5,47e-104	Length:	3011
Score:	1081.00	Matches:	199
Percent Similarity:	98.56%	Conservative:	6
Best Local Similarity:	95.67%	Mismatches:	3
Query Match:	93.59%	Indels:	0
DB:	20	Gaps:	0

US-09-899-303a-7 (1-633) x AAB59174 (1-3011)

```

QY 4 TTGGGTAAAGGTATGATGATACCTTACGTGCGGCTTCCGACCTCATGGGGTACATTCCG 63
DB 119 LeuGlyValAlaProLeuGlyAlaAlaArgAlaLeuAlaHisGlyValArgValLeu 158
QY 64 CTCGTGGGGGGCCCCCTAGGGGCTGCTGCAGAGCCCTGGCCATGGCGTCCGGGTTCTG 123
DB 139 LeuValGlyAlaProLeuGlyAlaAlaArgAlaLeuAlaHisGlyValArgValLeu 158
QY 124 GAAGAGCGGTGAACTATGCAACAGGGAATTGCTGCTTCTTCTATCTCTC 183
DB 159 GluAspGlyValAsnTrpAlaThrGlyAsnLeuProGlyCysSerPheSerIlePheLeu 178
QY 184 TTGGCTTAACTGTCGTGTGACCATTCAGCTTCGCTTATGAGTGCGCAAGCTGTCC 243
DB 179 LeuAlaLeuLeuSerCysLeuThrIleProAlaSerAlaTyrGlyValArgAsnValSer 198
QY 244 GGGATGATACATGTCAGAGACAGCTGTCGAACCTCAAGCATGTTGATGAGCGAGCGGAC 303
DB 199 GlyIleTrpHisValThrAsnAspCysSerAsnSerSerIleValTrpValAlaAlaAsp 218
QY 304 ATGATCATGCACACCCCGGGTGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 363
DB 219 ValIleMetHisThrProGlyCysValProCysValGlnGlnGlnGlnGlnGlnGlnGln 238
QY 364 TGGGTAGCGCTACACCCCGAGCTGCGAGTGAAGACGCGAGCGCTCCCACTAGACAAATA 423
DB 239 TrpValAlaLeuThrProThrLeuAlaAlaArgAsnAlaSerValProThrThrThrIle 258
QY 424 CGACGCGACGTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 483
DB 259 ArgArgHisValAspLeuLeuValGlyThrAlaAlaPheCysSerAlaMetValGly 278
QY 484 GATCTGCGGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 543
DB 279 AspLeuCysGlySerIlePheLeuValSerGlnLeuPheThrPheSerProAlaArgHis 298
QY 544 GAGACGCGTCAAGACTGCAATGCTCAATCTATCCGGCCACATAACAGGTACACGCTATG 603
DB 299 GluThrValGlnAspCysAsnSerIleTyrProGlyHisValSerGlnHisAlaMet 318
QY 604 GCTTGGATATGATGATGAACCTG 627
DB 319 AlaTrpAspMetMetMetAsnTrp 326

```

RESULT 13

AAB59174 standard; Protein; 3010 AA.

AC AAB59174;

DT 21-MAR-2001 (first entry)

DE Protein encoded by infectious Hepatitis C virus 1b genotype.

KM GBV-B; hepatitis C virus; HCV; vaccine.

OS Hepatitis C virus.

XX WO200075337-A1.

```

XX 14-DEC-2000.
PD 02-JUN-2000; 2000WO-US15293.
XX 04-JUN-1999; 99US-0137694.
PR (USSH ) US DEPT HEALTH & HUMAN SERVICES.
PA Bukh J, Yanagi M, Emerson SU, Purcell RH;
PI WPI; 2001-091214/10.
XX
DB New infectious nucleic acids of the GB virus-B clone, useful for
XX indirectly studying the molecular properties of hepatitis C virus (HCV)
XX and in developing vaccines and therapeutics for HCV
XX
PS Disclosure; Fig 7; 96pp; English.
XX
CC The present invention relates to GB virus-B. The nucleic acid molecules
XX of the invention are useful for indirectly studying the molecular
XX properties of hepatitis C virus (HCV). The infectious nucleic acid
XX sequence of the GB virus-B clone and the HCV/GBV-B chimeras may be used
XX in the development of vaccines and therapeutics for HCV.

```

XX Sequence 3010 AA:

Alignment Scores:

Pred. No.:	2.32e-103	Length:	3010
Score:	1075.00	Matches:	198
Percent Similarity:	98.08%	Conservative:	6
Best Local Similarity:	95.19%	Mismatches:	4
Query Match:	93.07%	Indels:	0
DB:	22	Gaps:	0

US-09-899-303a-7 (1-633) x AAB59174 (1-3010)

```

QY 4 TTGGGTAAAGGTATGATGATACCTTACGTGCGGCTTCCGACCTCATGGGGTACATTCCG 63
DB 119 LeuGlyValAlaProLeuGlyAlaAlaArgAlaLeuAlaHisGlyValArgValLeu 138
QY 64 CTCGTGGGGGGCCCCCTAGGGGCTGCTGCAGAGCCCTGGCCATGGCGTCCGGGTTCTG 123
DB 139 LeuValGlyAlaProLeuGlyAlaAlaArgAlaLeuAlaHisGlyValArgValLeu 158
QY 124 GAAGAGCGGTGAACTATGCAACAGGGAATTGCTGCTTCTTCTATCTCTC 183
DB 159 GluAspGlyValAsnTrpAlaThrGlyAsnLeuProGlyCysSerPheSerIlePheLeu 178
QY 184 TTGGCTTAACTGTCGTGTGACCATTCAGCTTCGCTTATGAGTGCGCAAGCTGTCC 243
DB 179 LeuAlaLeuLeuSerCysLeuThrIleProAlaSerAlaTyrGlyValArgAsnValSer 198
QY 244 GGGATGATACATGTCAGAGACAGCTGTCGAACCTCAAGCATGTTGATGAGCGAGCGGAC 303
DB 199 GlyIleTrpHisValThrAsnAspCysSerAsnSerSerIleValTrpValAlaAlaAsp 218
QY 304 ATGATCATGCACACCCCGGGTGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 363
DB 219 ValIleMetHisThrProGlyCysValProCysValGlnGlnGlnGlnGlnGlnGlnGln 238
QY 364 TGGGTAGCGCTACACCCCGAGCTGCGAGTGAAGACGCGAGCGCTCCCACTAGACAAATA 423
DB 239 TrpValAlaLeuThrProThrLeuAlaAlaArgAsnAlaSerValProThrThrThrIle 258
QY 424 CGACGCGACGTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 483
DB 259 ArgArgHisValAspLeuLeuValGlyThrAlaAlaPheCysSerAlaMetValGly 278
QY 484 GATCTGCGGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 543
DB 279 AspLeuCysGlySerIlePheLeuValSerGlnLeuPheThrPheSerProAlaArgHis 298

```

```

OY 544 GAGACGTCGACGACATTCATTCATCCGCCACATACAGTCACCGTATG 603
DB 299 GlnThrValGlnAspCysAsnCySerIleTyrProGlnHisValSerGlyHisArgMet 318
OY 604 GCTTGCGATATGATGATGAACTCG 627
DB 319 AlatrPaspMetMetMetAsnTirp 326

RESULT 14
AAR25854
ID AAR25854 standard; Protein; 416 AA.
XX
AC AAR25854;
XX
DT 21-JAN-1993 (first entry)
XX
DE HCV polypeptide 1.
XX
KM Recombinant vector; E. coli; diagnostic; reagent; type C hepatitis.
XX
OS Hepatitis C virus.
XX
PN JP04179482-A.
XX
PD 26-JUN-1992.
XX
PF 11-NOV-1990; 90JP-0304417.
XX
PR 11-NOV-1990; 90JP-0304417.
XX
PA (TOKU ) TOKUYAMA SODA KK.
XX
DR WPI; 1992-263663/32.
XX
N-PSDB; AAQ26981.
XX
PT Hepatitis C virus antigen expressed as recombinant in E.coli -
PT useful for diagnosis of hepatitis C virus infection
XX
PS Claim 1; Page 1-2; 66pp; Japanese.
XX
CC The sequences given in AAR25854-74 are hepatitis C virus proteins.
CC The genes encoding these proteins can each be used to prepare
CC recombinant vectors by ligating the gene of interest in to a vector
CC to be expressed in E. coli. These polypeptides are useful as
CC diagnostic reagents for type C hepatitis and they may be produced
CC efficiently by recombinant methods.
XX
SO Sequence 416 AA:

Alignment Scores:
Pred. No.: 1.88e-103 Length: 416
Score: 1073.00 Matches: 198
Percent Similarity: 98.56% Conservative: 7
Best Local Similarity: 95.19% Mismatches: 3
Query Match: 92.90% Indels: 0
DB: 13 Gaps: 0

US-09-899-303a-7 (1-633) x AAR25854 (1-416)
OY 4 TTGGGTAAAGTCATCGATACCTTACGTCGGCTTCCGCCACCTCATGGGTACATTCCG 63
DB 24 LeuGlyValIleAspTrpLeuThrCysGlyPheAlaAspLeuMetGlyTyrIlePro 43
OY 64 CTGCTGCGCCGCCCTAGGGGTCGTCACAGAGCCCTGGCGCATGGCGTCCGGGTTCTG 123
DB 44 LeuValGlyAlaProLeuGlyGlyAlaAlaArgAlaLeuAlaHisGlyValAlaGValLeu 63
OY 124 GAAGAGCGCGTGAACATGACAGAGGAATTCGCTGGTTCCTTCTATCTTCCTC 183
DB 64 GluAspIleValAsnTyrAlaThrGlyAsnLeuProGlySerPheSerIlePheLeu 83
OY 184 TTGGCTTATCTGTCCTGTCGACCATTCACAGCTTCGCTTATGAGTGGCCAAACGTCTC 243

```

```

DB 84 LeuAlaLeuSerCysLeuThrIleProAlaSerAlaTyrGluValAlaGAsnValSer 103
OY 244 GCGATGTACCAATGTACAGAAAGACTGCTCCAACTCAAGCAATGTGTATGAGCGACG 303
DB 104 GlyIleTyrHisValThrAsnAspCysSerAsnSerIleValTyrGluAlaAlaAsp 123
OY 304 ATGATCATGCACACACCCCGGGTGGCGCTTCGCTGGGAGAAACAATCTTCCGCTGC 363
DB 124 MetIleMetHisThrProGlyCysValProCysValArgGluSerAsnPheSerArgCys 143
OY 364 TTGGTACGCTACACCCCGAGCTCGCAGTACAGAACCCGACCGTCCCACTACGACAATA 423
DB 144 TrpAlaAlaLeuThrProThrLeuAlaAlaArgAsnSerIleProThrThrIle 163
OY 424 CGACGCCACGCTCATTTGCTGCTGGGGCGGCGCTTTCCTCCGCTATGTCAGTGGG 483
DB 164 ArgThrHisValAspLeuLeuValGlyAlaAlaAlaLeuCysSerAlaMetTyrValGly 183
OY 484 GATCTCTGCGGATCTGCTCTCTGCTCCAGCTGTTCACATCTCCGCTCGCGGCGAT 543
DB 184 AspLeuCysGlySerValPheLeuValSerGlnLeuPheThrPheSerProArgArgTyr 203
OY 544 GAGACGTCGACGACATTCATTCATTCATCCGCCACATACAGTCAACCGTATG 603
DB 204 GlnThrValGlnAspCysAsnCySerIleTyrProGlnHisValSerGlyHisArgMet 223
OY 604 GCTTGCGATATGATGATGAACTCG 627
DB 224 AlatrPaspMetMetMetAsnTirp 231

RESULT 15
AAR25876
ID AAR25876 standard; Protein; 445 AA.
XX
AC AAR25876;
XX
DT 21-JAN-1993 (first entry)
XX
DE HK1.
XX
KM Recombinant vector; E. coli; diagnostic; reagent; type C hepatitis.
XX
OS Hepatitis C virus.
XX
FH Key Location/Qualifiers
FT protein 25..441
FT /*tag= "a
FT /*note= "Sequence AAR25854"

JP04179482-A.
PN
PD 26-JUN-1992.
XX
PF 11-NOV-1990; 90JP-0304417.
XX
PR 11-NOV-1990; 90JP-0304417.
XX
PA (TOKU ) TOKUYAMA SODA KK.
XX
DR WPI; 1992-263663/32.
XX
N-PSDB; AAQ27003.
XX
PT Hepatitis C virus antigen expressed as recombinant in E.coli -
PT useful for diagnosis of hepatitis C virus infection
XX
PS Disclosure; Fig 2; 66pp; Japanese.
XX
CC The sequences given in AAR25876-95 are encoded by the claimed hepatitis
CC C virus genes of the invention which have been inserted into an E.
CC coli vector. These polypeptides are useful as diagnostic reagents
CC for type C hepatitis and they may be produced efficiently by
CC recombinant DNA techniques.
XX

```

SQ Sequence 445 AA:

## Alignment Scores:

Pre. No.:	1.93e-103	Length:	445
Score:	1073.00	Matches:	198
Percent Similarity:	98.56%	Conservative:	7
Best Local Similarity:	95.19%	Mismatches:	3
Query Match:	92.90%	Indels:	0
DB:	13	Gaps:	0

US-09-899-303A-7 (1-633) x AAR25876 (1-445)

```
OY      4  TTGGGTAAGCTATGATGATCCCTTACGTGCGGCTTCGCCGACCTCATGGGGTACATTCGG  63
      |||
DB     49  LeuGlyLyValIleAspThrLeuThrCysGlyPheAlaAspLeuMetGlyTyrIlePro  68
OY     64  CTCGTGCGGCGCCCTTACGGGGTGCCTGCCACAGCCCTGGCCATGAGCGTCCGGGTTCTG  123
      |||
DB     69  LeuValGlyAlaProLeuGlyGlyAlaAlaArgAlaLeuAlaHisGlyValArgValLeu  88
OY     124 GAAGACGGCGGTGAACATGACCAACAGGGAATTTGCTGTGCTTTCTATCTTCTC  183
      |||
DB     89  GluAspGlyValAsnTyrAlaThrGlyAsnLeuProGlyCysSerPheSerIlePheLeu  108
OY     184 TTGGGTTTACTGTCCCTGTGTGACCAATCCAGTTCGCTTATGAGTGGCAACGTGCC  243
      |||
DB     109 LeuAlaLeuLeuSerCysLeuThrIleProAlaSerAlaTyrGlyValArgAsnValSer  128
OY     244 GGGATGTACCATGTATCGAAGACAGTGTCCCACTCAAGCATTTGTATGAGGACGGGAC  303
      |||
DB     129 GlyIleTyrHisValIleThrAsnAspCysSerAsnSerSerIleValIyrGlyAlaAlaAsp  148
OY     304 ATGATCATGACACCCCGCGGTGCGTGCCTTCGTTCCGGAGAACAACTTTCCCGCTGC  363
      |||
DB     149 MetIleMetHisThrProGlyCysValProCysValArgGlySerAsnPheSerArgCys  168
OY     364 TGGGTAGCCCTCACCCCGCGCTCGAGCTAGAGACGCCAGCGTCCCGCACTACGACATA  423
      |||
DB     169 TrpValAlaLeuThrProThrLeuAlaAlaArgAsnSerSerIleProThrThrIle  188
OY     424 CGACGCCACGTGATTTGCTGCTGGGGGGGCTGTTCTGTCCGCTATGTACGTGGGG  483
      |||
DB     189 ArgAlaGlyHisValAspLeuLeuValGlyAlaAlaAlaLeuCysSerAlaMetTyrValGly  208
OY     484 GATCTCTGGGATCTGTCTTCTCTGCTGCCACGCTGTTACCACTTCGGCTCGCGGCAT  543
      |||
DB     209 AspLeuCysGlySerValPheLeuValSerGlnLeuPheThrPheSerProArgArgTyr  228
OY     544 GAGACGGTGCAGAGCTGCAATGCTAATCTATCCGGCCACATAACAGAGTCAACGCTATG  603
      |||
DB     229 GluThrValGlnAspCysAsnCysSerIleTyrProGlyHisValSerGlyHisArgMet  248
OY     604 GCTTGGGATATGATGATGAAGTGG  627
      |||
DB     249 AlaTrpAspMetMetMetAsnTrp  256
```

Search completed: February 12, 2003, 08:58:46  
Job time : 51 secs

GenCore version 5.1.3  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - protein search, using frame\_plus\_n2p model

Run on: February 12, 2003, 08:54:59 ; Search time 14 Seconds  
(without alignments)  
2660.673 Million cell updates/sec

Title: US-09-899-303A-7

Perfect score: 1155  
Sequence: 1 ATGTGTGTAAGTCATCGA.....TGATGATGAAGTGTATAG 633

Scoring table:

BLOSUM62	Xgapext 0.5
Ygapop 10.0	Ygapext 0.5
Fgapop 6.0	Fgapext 7.0
Delop 6.0	Delext 7.0

Searched: 262574 segs, 29422922 residues

Total number of hits satisfying chosen parameters: 525148

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:  
-MODEL=frame+ n2p.model -DEV=xip  
-O/cgn2.1/USPTO.spool/US09899303/r/unat.12022003.085450.28314/app.query.fasta.1.775  
-DB=Issued Patents\_AA -QFMT=fasta -SUFFIX=n2p.ra1 -MTNMATCH=0.1 -LOOPEL=0  
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi  
-LIST=45 -DOCLIN=200 -THR.SCORE=pct -THR.MAX=100 -THR.MIN=0 -ALIGN=15  
-MODE=LOCAL -OUTFMT=ptc -NRM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000  
-USER=US09899303 @cgn1.1.12 @unat.12022003.085450.28314 -NCPU=6 -ICPU=3  
-NO.XLPHY -NO.MMAP -LARGEQUERY -NEG.SCORES=0 -WAIT -LONGLOG -DEV.TIMEOUT=120  
-MARN.TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -YGAPOP=6 -YGAPEXT=7  
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database:

Issued Patents\_AA:\*  
1: /cgn2.6/prodata/1/iaa/5A.COMB.pep:\*  
2: /cgn2.6/prodata/1/iaa/5B.COMB.pep:\*  
3: /cgn2.6/prodata/1/iaa/6A.COMB.pep:\*  
4: /cgn2.6/prodata/1/iaa/6B.COMB.pep:\*  
5: /cgn2.6/prodata/1/iaa/6C.COMB.pep:\*  
6: /cgn2.6/prodata/1/iaa/backfile1.pep:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length DB	ID	Description
1	1118	96.8	209	4	US-08-612-973-8
2	1118	96.8	209	4	US-08-927-597-8
3	1113	96.4	263	4	US-08-612-973-6
4	1113	96.4	263	4	US-08-927-597-6
5	1109	96.0	210	4	US-08-612-973-14
6	1109	96.0	210	4	US-08-927-597-14
7	1108	95.9	632	4	US-08-612-973-48
8	1108	95.9	632	4	US-08-927-597-48
9	1108	95.9	809	4	US-08-612-973-50
10	1108	95.9	809	4	US-08-927-597-50
11	1084	93.9	340	1	US-08-462-195-2
12	1084	93.9	340	2	US-08-636-883-2

13	1084	93.9	340	3	US-09-127-829-2
14	1081	93.6	3010	4	US-09-014-416-3
15	1081	93.6	3011	4	US-09-014-416-5
16	1065	92.2	340	1	US-08-462-195-4
17	1065	92.2	340	2	US-08-636-883-4
18	1065	92.2	340	3	US-09-127-829-4
19	1060	91.8	389	1	US-08-324-977-4
20	1060	91.8	389	2	US-08-324-977-10
21	1060	91.8	389	2	US-08-384-616-4
22	1060	91.8	389	2	US-08-384-616-10
23	1060	91.8	389	2	US-08-904-686A-4
24	1060	91.8	389	2	US-08-904-686A-10
25	1060	91.8	389	4	US-09-315-850-4
26	1060	91.8	389	4	US-09-315-850-12
27	1060	91.8	2013	1	US-08-324-977-12
28	1060	91.8	2013	2	US-08-384-616-12
29	1060	91.8	2013	2	US-08-904-686A-12
30	1060	91.8	2013	4	US-09-315-850-12
31	1060	91.8	3010	1	US-08-324-977-2
32	1060	91.8	3010	1	US-08-324-977-14
33	1060	91.8	3010	2	US-08-384-616-2
34	1060	91.8	3010	2	US-08-384-616-14
35	1060	91.8	3010	2	US-08-904-686A-2
36	1060	91.8	3010	2	US-08-904-686A-14
37	1060	91.8	3010	4	US-09-315-850-2
38	1060	91.8	3010	4	US-09-315-850-14
39	1006	87.1	967	1	US-08-188-281B-13
40	1006	87.1	967	5	PCT-US94-07280-13
41	1006	87.1	967	5	PCT-US95-01087-13
42	1006	87.1	1648	1	US-08-188-281B-12
43	1006	87.1	1648	5	PCT-US94-07280-12
44	1006	87.1	1648	5	PCT-US95-01087-12
45	1006	87.1	3011	1	US-08-188-281B-1

#### ALIGNMENTS

RESULT 1  
US-08-612-973-8  
Sequence 8, Application US/08612973  
Patent No. 6150134

GENERAL INFORMATION:  
APPLICANT: MAERTENS, GEERT  
APPLICANT: BOSMAN, FONS  
APPLICANT: DE MARTYNOFF, GUY  
APPLICANT: BUYSSE, MARIE-ANGE  
TITLE OF INVENTION: PURIFIED HEPATITIS C VIRUS ENVELOPE  
TITLE OF INVENTION: PROTEINS FOR DIAGNOSTIC AND THERAPEUTIC USE  
NUMBER OF SEQUENCES: 111  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: NIXON & VANDERHAYE P.C.  
STREET: 1100 NORTH GLEBE ROAD  
CITY: ARLINGTON  
STATE: VIRGINIA  
COUNTRY: U.S.A.  
ZIP: 22201-4714

MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/612,973  
FILING DATE: 11-MAR-1996  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: BYRNE, THOMAS E.  
REGISTRATION NUMBER: 32,205  
REFERENCE/DOCKET NUMBER: 1487-10  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (703) 816-4000  
TELEFAX: (703) 816-4100  
INFORMATION FOR SEQ ID NO: 8:

```

SEQUENCE CHARACTERISTICS:
; LENGTH: 209 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-612-973-8

Alignment Scores:
Pred. No.: 8.35e-111 Length: 209
Score: 1118.00 Matches: 209
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 96.80% Indels: 0
DB: 4 Gaps: 0

US-09-899-303a-7 (1-633) x US-08-612-973-8 (1-209)
QY 1 ATGTTGGTAAAGTCAATGATACCTTACGTGGCGCTGCGCAGCTCAATGGGTACATT 60
DB 1 MetLeuGlyLysValIleAspThrLeuThrCysGlyPheAlaAspLeuMetGlyTyrIle 20
QY 61 CGGCTGTGGGGGCCCCCTAGGGGGTGCCTCCAGAGCCCTGGCGATGGCGTCCGGATT 120
DB 21 ProlLeuValIGlyAlaProlLeuGlyGlyAlaAlaArgAlaLeuAlaHisGlyValArgVal 40
QY 121 CTGGGAAGAGGGGTGAATATGCAAGAGGAATTTCCCTGGTTCCTCTTCTATCTTC 180
DB 41 LeuGluAspGlyValAsnTyrAlaThrGlyAsnLeuProlGlyCysSerPheSerIlePhe 60
QY 181 CTTTGGGCTTTACTGTCCTGTGACCATTTCCAGCTTCCGCTTATGAGTGGCCAGCGT 240
DB 61 LeuLeuAlaLeuLeuSerCysLeuThrIleProlAlaSerAlaTyrGluValArgAsnVal 80
QY 241 TCCGGGATGTACCATGTGCAAGAGCACTGCTCCAACTCAAGCATTTGTATGAGCGAGG 300
DB 81 SerGlyMetTyrHisValIleThrAsnAspCysSerAsnSerIleValIlyrGluAlaAla 100
QY 301 GACATGATATGACACACCCCCGGGTGCGCTGCGCTGCGGAGAACAACTTTCCCGG 360
DB 101 AspMetIleMetHisThrProlGlyCysValProCysValArgGluAsnAsnSerSerArg 120
QY 361 TGCCTGGGTAGCGCTCACCCCCAGCTCGAGCTAGGAGACGCCAGCTCCCACTAGACAG 420
DB 121 CysTrpValAlaLeuThrProThrIleuAlaAlaAlaArgAsnAlaSerValProThrThr 140
QY 421 ATACAGACGCCAGTGCATTTGCTGCTGGGGGCGCTGCTTGTCTGCTACATGACGTG 480
DB 141 IleArgGlnHisValAspLeuLeuValIGlyAlaAlaAlaPheCysSerAlaMetGlyVal 160
QY 481 GGGGATCTCTGGGATCTCTTCTCCGTCGCTCCAGCTGTACCATGTCGCTCGCGGG 540
DB 161 GlysAspLeuCysGlySerValPheLeuValSerGlnLeuPheThrIleSerProArgArg 180
QY 541 CATGAGACGGCTCAGACATGCAATGCTCAATCTATCCGGGCACATACAGGTACCCGT 600
DB 181 HisGluThrValIGlnAspCysAsnCysSerIleTyrProlGlnHisIleThrGlnHisArg 200
QY 601 ATGCTTGGGATATGATGATGAATCG 627
DB 201 MetAlaTrpAspMetMetMetAsnTrp 209

RESULT 2
US-08-927-597-8
; Sequence 8, Application US/08927597
; Patent No. 6245503
; GENERAL INFORMATION:
; APPLICANT: NAERTENS, GEERT
; APPLICANT: BOSMAN, FONS
; APPLICANT: DE MARTYNOFF, GUY
; APPLICANT: BUYSSE, MARIE-ANGE
; TITLE OF INVENTION: PURIFIED HEPATITIS C VIRUS ENVELOPE
; TITLE OF INVENTION: PROTEINS FOR DIAGNOSTIC AND THERAPEUTIC USE
; NUMBER OF SEQUENCES: 111

```

```

CORRESPONDENCE ADDRESS:
; ADDRESSEE: NIXON & VANDERHYE P. C.
; STREET: 1100 NORTH GLEBE ROAD
; CITY: ARLINGTON
; STATE: VIRGINIA
; COUNTRY: U.S.A.
; ZIP: 22201-4714
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/927,597
; FILING DATE:
; CLASSIFICATION:
; PRIORITY APPLICATION DATA:
; APPLICATION NUMBER: US 08/612,973
; FILING DATE: 11-MAR-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: BYRNE, THOMAS E.
; REGISTRATION NUMBER: 32,205
; REFERENCE/DOCKET NUMBER: 1487-10
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 816-4000
; TELEFAX: (703) 816-4100
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 209 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-927-597-8

Alignment Scores:
Pred. No.: 8.35e-111 Length: 209
Score: 1118.00 Matches: 209
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 96.80% Indels: 0
DB: 4 Gaps: 0

US-09-899-303a-7 (1-633) x US-08-927-597-8 (1-209)
QY 1 ATGTTGGTAAAGTCAATGATACCTTACGTGGCGCTGCGCAGCTCAATGGGTACATT 60
DB 1 MetLeuGlyLysValIleAspThrLeuThrCysGlyPheAlaAspLeuMetGlyTyrIle 20
QY 61 CGGCTGTGGGGGCCCCCTAGGGGGTGCCTCCAGAGCCCTGGCGATGGCGTCCGGATT 120
DB 21 ProlLeuValIGlyAlaProlLeuGlyGlyAlaAlaArgAlaLeuAlaHisGlyValArgVal 40
QY 121 CTGGGAAGAGGGGTGAATATGCAAGAGGAATTTCCCTGGTTCCTCTTCTATCTTC 180
DB 41 LeuGluAspGlyValAsnTyrAlaThrGlyAsnLeuProlGlyCysSerPheSerIlePhe 60
QY 181 CTTTGGGCTTTACTGTCCTGTGACCATTTCCAGCTTCCGCTTATGAGTGGCCAGCGT 240
DB 61 LeuLeuAlaLeuLeuSerCysLeuThrIleProlAlaSerAlaTyrGluValArgAsnVal 80
QY 241 TCCGGGATGTACCATGTGCAAGAGCACTGCTCCAACTCAAGCATTTGTATGAGCGAGG 300
DB 81 SerGlyMetTyrHisValIleThrAsnAspCysSerAsnSerIleValIlyrGluAlaAla 100
QY 301 GACATGATATGACACACCCCCGGGTGCGCTGCGCTGCGGAGAACAACTTTCCCGG 360
DB 101 AspMetIleMetHisThrProlGlyCysValProCysValArgGluAsnAsnSerSerArg 120
QY 361 TGCCTGGGTAGCGCTCACCCCCAGCTCGAGCTAGGAGACGCCAGCTCCCACTAGACAG 420
DB 121 CysTrpValAlaLeuThrProThrIleuAlaAlaAlaArgAsnAlaSerValProThrThr 140
QY 421 ATACAGACGCCAGTGCATTTGCTGCTGGGGGCGCTGCTTGTCTGCTACATGACGTG 480

```

Db 141 ILeArGrHsValAspLeuValGlyAlaAlaPheCysSerAlaMetLysVal 160  
QY 481 GGGATGTCGCGGATGTCCTTCCTGTCGCCAGCTGTCACCATCTGCGCCGG 540  
Db 161 GlyAspLeuGlySerValPheLeuValSerGlnLeuPheThrIleSerProArgArg 180  
QY 541 CATGAGACGGTGACGAGCTGCAATGCTCAATCTATCCCGGCCACATTAACAGTCCACCGT 600  
Db 181 HisGlnThrValGlnAspCysAsnGlySerIleLysTrpGlyHisIleThrGlyHisArg 200  
QY 601 ATGCGTTGGGATATGATGATGACTGG 627  
Db 201 MetAlaTrpAspMetMetMetLsnTrp 209  
RESULT 3  
US-08-612-973-6  
; Sequence 6, Application US/08612973  
; Patent No. 6130134  
; GENERAL INFORMATION:  
; APPLICANT: MAERTENS, GEERT  
; APPLICANT: BOSMAN, FONS  
; APPLICANT: DE MARTYNOFF, GUY  
; APPLICANT: BUYSSE, MARIE-ANGE  
; TITLE OF INVENTION: PURIFIED HEPATITIS C VIRUS ENVELOPE  
; TITLE OF INVENTION: PROTEINS FOR DIAGNOSTIC AND THERAPEUTIC USE  
; NUMBER OF SEQUENCES: 111  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: NIXON & VANDERHAYE P.C.  
; STREET: 1100 NORTH GLEBE ROAD  
; CITY: ARLINGTON  
; STATE: VIRGINIA  
; COUNTRY: U.S.A.  
; ZIP: 22201-4714  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/612,973  
; FILING DATE: 11-MAR-1996  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: BYRNE, THOMAS E.  
; REGISTRATION NUMBER: 32,205  
; REFERENCE/DOCKET NUMBER: 1487-10  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (703) 816-4000  
; TELEFAX: (703) 816-4100  
; INFORMATION FOR SEQ ID NO: 6:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 263 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; US-08-612-973-6  
Alignment Scores:  
Pred. No.: 3,13e-110 Length: 263  
Score: 1113.00 Matches: 207  
Percent Similarity: 100.00% Conservative: 2  
Best Local Similarity: 99.04% Mismatches: 0  
Query Match: 96.36% Indels: 0  
DB: 4 Gaps: 0  
US-09-899-303a-7 (1-633) x US-08-612-973-6 (1-263)  
QY 1 ATGTTGGGTAAAGTCAATGATACCTTACGTCGCGCTTCCGCGACCTCATGGGGTACATT 60  
Db 1 MetLeuGlyGlyValIleAspThrLeuThrCysGlyPheAlaAspLeuValGlyTyrIle 20  
QY 61 CCGCTCGTCGGCGCCCTTAAGGGGTGCTGCCAGAGCCCTGGCGCATGGCGTCCGGGTT 120

Db 21 ProLeuValGlyAlaProLeuGlyGlyAlaAlaArgAlaLeuAlaHisGlyValArgVal 40  
QY 121 CTGGAGACGGCGTGAAGCTATGCAACAGGGGAATTGGCTGGTCTCTTCTCATCTTC 180  
Db 41 LeuGlnAspGlyValAsnTrpAlaThrIleGlyAsnLeuProGlyCysSerPheSerIlePhe 60  
QY 181 CTCTTGGCTTACTGTCTCTGTCTGACCATTCACACTTCGCTTATGAGAGTCCGACAGTG 240  
Db 61 LeuLeuAlaLeuLeuSerCysLeuThrValProAlaSerAlaTyrGluValArgAsnVal 80  
QY 241 TCCGGGATGTAACATGTCACGACGAGCTGTCCAACTCAACATGATGTATAGACGACGG 300  
Db 81 SerGlyMetLysHisValThrAsnAspCysSerAsnSerIleValTyrGluAlaAla 100  
QY 301 GACATGATCATGACACACCCCGGGGTGGCTGCCCTGGCGGAGAACAACTTCTCCCGC 360  
Db 101 AspMetIleMetHisThrProGlyCysValProCysValArgGlnAsnAsnSerSerArg 120  
QY 361 TGCTGGGTAGCGCTACACCCCGCCAGCTGACAGTGAAGCGCACGTCCTACAGACA 420  
Db 121 CysTrpValAlaLeuThrProThrLeuAlaAlaArgAsnAlaSerValProThrThrThr 140  
QY 421 ATACGAGCCGACGTGATTTGCTGTTGGGGCGGCTGCTTCTGTCTATGTATAGCTG 480  
Db 141 ILeArGrHsValAspLeuValGlyAlaAlaPheCysSerAlaMetLysVal 160  
QY 481 GGGATGTCGCGGATGTCCTTCCTGTCGCCAGCTGTCACCATCTGCGCCGGCGCG 540  
Db 161 GlyAspLeuGlySerValPheLeuValSerGlnLeuPheThrIleSerProArgArg 180  
QY 541 CATGAGACGGTGACGAGCTGCAATGCTCAATCTATCCCGGCCACATTAACAGTCCACCGT 600  
Db 181 HisGlnThrValGlnAspCysAsnGlySerIleLysTrpGlyHisIleThrGlyHisArg 200  
QY 601 ATGCGTTGGGATATGATGATGACTGG 627  
Db 201 MetAlaTrpAspMetMetMetLsnTrp 209  
RESULT 4  
US-08-927-597-6  
; Sequence 6, Application US/08927597  
; Patent No. 6245503  
; GENERAL INFORMATION:  
; APPLICANT: MAERTENS, GEERT  
; APPLICANT: BOSMAN, FONS  
; APPLICANT: DE MARTYNOFF, GUY  
; APPLICANT: BUYSSE, MARIE-ANGE  
; TITLE OF INVENTION: PURIFIED HEPATITIS C VIRUS ENVELOPE  
; TITLE OF INVENTION: PROTEINS FOR DIAGNOSTIC AND THERAPEUTIC USE  
; NUMBER OF SEQUENCES: 111  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: NIXON & VANDERHAYE P.C.  
; STREET: 1100 NORTH GLEBE ROAD  
; CITY: ARLINGTON  
; STATE: VIRGINIA  
; COUNTRY: U.S.A.  
; ZIP: 22201-4714  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/927,597  
; FILING DATE:  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/612,973  
; FILING DATE: 11-MAR-1996  
; ATTORNEY/AGENT INFORMATION:  
; NAME: BYRNE, THOMAS E.  
; REGISTRATION NUMBER: 32,205

```

: REFERENCE/DOCKET NUMBER: 1487-10
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (703) 816-4000
: TELEFAX: (703) 816-4100
: INFORMATION FOR SEQ ID NO: 6:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 263 amino acids
: TYPE: amino acid
: TOPOLOGY: linear
: MOLECULE TYPE: protein
: US-08-927-597-6

Alignment Scores:
Pred. No.: 3,13e-110      Length: 263
Score: 1113.00           Matches: 207
Percent Similarity: 100.00%      Conservative: 2
Best Local Similarity: 99.04%     Mismatches: 0
Query Match: 96.36%              Indels: 0
DB: 4                          Gaps: 0

US-09-899-303A-7 (1-633) x US-08-927-597-6 (1-263)

QY 1 ATGTTGGGTAAGGTCATCGATGATACCCCTTACGCGGGCTTCCGCGACCTCATGGGGTACATT 60
Db 1 MetLeuGlyValIleAspThrLeuThnCysGlyPheAlaSerLeuValGlyTyrIle 20

QY 61 CCGCTCGTGGCGCCCCCTTAGGGGGGTGCTCCAGACCCCTGGCGGACATGGCGTCCGGGTT 120
Db 21 ProLeuValGlyAlaProLeuGlyGlyAlaAlaIArgAlaLeuAlaHisGlyValArgVal 40

QY 121 CTGGAAAGCGCGGACATATGCAAGAGGAAATTGGCTGGTTCCTCTCATCTTC 180
Db 41 LeuGluAspGlyValAsnTyrAlaThrGlyAsnLeuProGlyCysSerPheSerIlePhe 60

QY 181 CTCTTGAGCTTACTGCTGCTCTGTGACCATTCACGCTTCCGCTTATGAGGTGGCCAAAGT 240
Db 61 LeuLeuAlaLeuLeuSerCysLeuThrValProAlaSerIleTyrGlyValAlaArgAsnVal 80

QY 241 TTCGGGAGTATGCATATGTCAGAACAGCACTGCTCCAACTCAAGACATTTGTATAGAGGAGCG 300
Db 81 SerGlyMetTyrHisValThrAsnAspCysSerAsnSerIleValTyrGluAlaIala 100

QY 301 GACATGATCATGACGACACCCCGGGGTGGCTGCCCTTGGGAGAACAACTCTTCCCGC 360
Db 101 AspMetIleMetHisThrProGlyCysValProCysValArgGluAsnAsnSerSerArg 120

QY 361 TGTCTGGTAGGCGTCACCCCGACGCTGCGAGTAGAAGAACCCAGCGCCCTACATGACACA 420
Db 121 CysThrValAlaLeuThrProThrLeuAlaIalaArgAsnAlaSerValProThrThrThr 140

QY 421 ATACGACGCCACGTCGATTGTGCTGTTGGGGGGCTGCTTTTGTTCCGCTATGACGTG 480
Db 141 IleArgIArgHisValAspLeuIleValGlyAlaAlaIaIaPheCysSerIaIaMetTyrVal 160

QY 481 GGGGATCTCTCGGATATCTGTCCTTCTCTGTCCTCCAGCTGTTCACATCTCTGCTGCGCGG 540
Db 161 GlyAspLeuCysGlySerValPheLeuValSerGluLeuPheThrIleSerProAlaArg 180

QY 541 CATAGAGCGGTGCGAGCATGTCGAATTTGTCATTCATCCCGGACACATTAAGGTGACCGG 600
Db 181 HisGluThrValGlnAspCysAsnGlySerIleTyrProGlyHisIleThrGlyHisArg 200

QY 601 ATGGCTTGGGATATGATGATGAATCGG 627
Db 201 MetaIaIArgAspMetMetMetAsnTrrp 209

RESULT 5
US-08-612-973-14
: Sequence 14, Application US/08612973
: Patent No. 6150134
: GENERAL INFORMATION:
: APPLICANT: MAERTENS, GEERT
: APPLICANT: BOSMAN, FONS

```

QY	421	ATACGAGCCACACGTCGATTTGGTGGGGGGGCTCTTTGGTCCGCTATGTAAGTGC	480
QY	422		
Db	141	ITLkgrghrhstValAspLeuValGlyAlaAlaAlaPheCysSerAlaMetTyrVal	150
QY	481	GGGAGATCTGCGGGATCTGTCTTCCTCGCTCCAGCTGTTCACCATCTCGCTCGCGG	540
Db	161	GlyAspLeuCysGlySerValPheLeuValSerGlnLeuPheThrIleSerProArgArg	180
QY	541	CATGAGACGGTGCACGAGACTGCAATTGGCTCAATCTATCCCGGCCACATACAGGTCAAGT	600
Db	181	HISgluThrValGlnAspCysAsnCysSerIleTyrProGlnHisIleThrGlnHisArg	200
QY	601	ATGGCTGGGATATGATGATGATCAACTGG	627
Db	201	MetAlaTrpAspMetMetMetAsnTrp	209

NAME: BYRNE, THOMAS E.  
REGISTRATION NUMBER: 32,205  
REFERENCE/DOCKET NUMBER: 1487-10  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (703) 816-4000  
TELEFAX: (703) 816-4100  
INFORMATION FOR SEQ. ID NO: 48:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 692 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-612-973-48

Alignment Scores:  
Pred. No.: 1.59e-109 Length: 692  
Score: 1108.00 Matches: 206  
Percent Similarity: 100.00% Conservative: 2  
Best Local Similarity: 99.04% Mismatches: 0  
Query Match: 95.93% Indels: 0  
DB: 4 Gaps: 0

US-09-899-303A-7 (1-633) x US-08-612-973-48 (1-692)

QY 4 TTGGGTAAGGTCAATCCATACCTTACGTGGCGCTTCCGCGCACTCATGGGTACATTCCG 63  
Db 2 LeuG1yLysVal11IleAspThrLeuThrCysG1yPheAlaAspLeuVal1G1yTyr1IlePro 21  
QY 64 CTCGTGGGCGCCCGCCAGGAGGGGCGCCAGAGCCCTGGCGCATGGCGGTCGGGTCTG 123  
Db 22 LeuVal1G1yAlaProLeuG1yG1yAlaAlaArgAlaLeuAlaHisG1yValAlaGValLeu 41  
QY 124 GAAGACGGCGGTGAATGACAGACAGGGAATTGCGTGGTCTCTCTTCTATCTTCCTC 183  
Db 42 GluAspG1yValAlaAsnTyrAlaThrG1yAlaAsnLeuProG1ySerpPheSer1IlePheLeu 61  
QY 184 TTGGCTTTACTGCTCTGTCGACCAATTCGAGTTCGCTTATGAGGCGCCAGAGTTC 243  
Db 62 LeuAlaLeuLeuSerCysLeuThrValProAlaSerAlaTyrG1yValAlaGAsnValSer 81  
QY 244 GGGATGTACCATGTCAAGACGACTGCTCCAGCTCAAGCAAGATTGATGAGGAGCGGAC 303  
Db 82 GlyMetTyrHisVal1ThrAsnAspCysSerAsnSerSer1IleVal1yTyrG1yAlaAlaAsp 101  
QY 304 ATGATCATGCACACCCCGGGTGCCTGCGCTTCCGAGAGACAATCTTCCGCTGC 363  
Db 102 Met1IleMetHisThrProG1yCysValProCysValAlaG1yAlaG1yAsnAsnSerSerArgCys 121  
QY 364 TGGGTAGCCGCTACACCCCGGCTCGAGCTCGAGCTAGAGAGCCGAGCTCCCACTAGACATA 423  
Db 122 TrrValAlaLeuThrProThrLeuAlaAlaArgAlaSerAlaProThrThrThrIle 141  
QY 424 CGAGCGACGATGATTTGCTGCTGGGGGCGGTCTTCTGTTCCGCTATGATGAGTGGG 483  
Db 142 ArgArgHisValAlaAspLeuLeuVal1G1yAlaAlaAlaPheCysSerAlaMetTyrValG1y 161  
QY 484 GATCTGCGGATCTGTCTTCTCTGCTGCTCCAGCTGTTCACATCTGCGCTCCCGGAT 543  
Db 162 AspLeuCysG1ySerValPheLeuValSerGlnLeuPheThr1IleSerProAlaGlnHis 181  
QY 544 GAAGACGGGTGAATGACAGACGAGGAATTGCTGCTGCTTCTTCTTCTATCTTCCTC 603  
Db 182 GluThrValG1yAlaAspCysAsnCysSer1IleTyrProG1yHis1IleThrG1yHisArgMet 201  
QY 604 GCTGGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 627  
Db 202 AlaTrrAspMetMetMetMetMetMetMetMetMetMetMetMetMetMetMetMetMet 209

RESULT 8  
US-08-927-597-48  
Sequence 48, Application US/08927597  
Patent No 6245503  
GENERAL INFORMATION:

APPLICANT: MAERTENS, GEERT  
APPLICANT: BOSMAN, FONS  
APPLICANT: DE MARTYNOFF, GUY  
APPLICANT: BUOYE, MARIE-ANGE  
TITLE OF INVENTION: PURIFIED HEPATITIS C VIRUS ENVELOPE  
TITLE OF INVENTION: PROTEINS FOR DIAGNOSTIC AND THERAPEUTIC USE  
NUMBER OF SEQUENCES: 111  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: NIXON & VANDERHUYE P.C.  
STREET: 1100 NORTH GLEBE ROAD  
CITY: ARLINGTON  
STATE: VIRGINIA  
COUNTRY: U.S.A.  
ZIP: 22201-4714  
COMPUTER READABLE FORM:  
MEDIUM TYPE: floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/927,597  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/612,973  
FILING DATE: 11-MAR-1996  
ATTORNEY/AGENT INFORMATION:  
NAME: BYRNE, THOMAS E.  
REGISTRATION NUMBER: 32,205  
REFERENCE/DOCKET NUMBER: 1487-10  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (703) 816-4000  
TELEFAX: (703) 816-4100  
INFORMATION FOR SEQ. ID NO: 48:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 692 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-927-597-48

Alignment Scores:  
Pred. No.: 1.59e-109 Length: 692  
Score: 1108.00 Matches: 206  
Percent Similarity: 100.00% Conservative: 2  
Best Local Similarity: 99.04% Mismatches: 0  
Query Match: 95.93% Indels: 0  
DB: 4 Gaps: 0

US-09-899-303A-7 (1-633) x US-08-927-597-48 (1-692)

QY 4 TTGGGTAAGGTCAATCCATACCTTACGTGGCGCTTCCGCGCACTCATGGGTACATTCCG 63  
Db 2 LeuG1yLysVal11IleAspThrLeuThrCysG1yPheAlaAspLeuVal1G1yTyr1IlePro 21  
QY 64 CTCGTGGGCGCCCGCCAGGAGGGGCGCCAGAGCCCTGGCGCATGGCGGTCGGGTCTG 123  
Db 22 LeuVal1G1yAlaProLeuG1yG1yAlaAlaArgAlaLeuAlaHisG1yValAlaGValLeu 41  
QY 124 GAAGACGGCGGTGAATGACAGACAGGGAATTGCGTGGTCTCTCTTCTATCTTCCTC 183  
Db 42 GluAspG1yValAlaAsnTyrAlaThrG1yAlaAsnLeuProG1ySerpPheSer1IlePheLeu 61  
QY 184 TTGGCTTTACTGCTCTGTCGACCAATTCGAGTTCGCTTATGAGGCGCCAGAGTTC 243  
Db 62 LeuAlaLeuLeuSerCysLeuThrValProAlaSerAlaTyrG1yValAlaGAsnValSer 81  
QY 244 GGGATGTACCATGTCAAGACGACTGCTCCAGCTCAAGCAAGATTGATGAGGAGCGGAC 303  
Db 82 GlyMetTyrHisVal1ThrAsnAspCysSerAsnSerSer1IleVal1yTyrG1yAlaAlaAsp 101  
QY 304 ATGATCATGCACACCCCGGGTGCCTGCGCTTCCGAGAGACAATCTTCCGCTGC 363  
Db 102 Met1IleMetHisThrProG1yCysValProCysValAlaG1yAlaG1yAsnAsnSerSerArgCys 121  
QY 364 TGGGTAGCCGCTACACCCCGGCTCGAGCTCGAGCTAGAGAGCCGAGCTCCCACTAGACATA 423  
Db 122 TrrValAlaLeuThrProThrLeuAlaAlaArgAlaSerAlaProThrThrThrIle 141  
QY 424 CGAGCGACGATGATTTGCTGCTGGGGGCGGTCTTCTGTTCCGCTATGATGAGTGGG 483  
Db 142 ArgArgHisValAlaAspLeuLeuVal1G1yAlaAlaAlaPheCysSerAlaMetTyrValG1y 161  
QY 484 GATCTGCGGATCTGTCTTCTCTGCTGCTCCAGCTGTTCACATCTGCGCTCCCGGAT 543  
Db 162 AspLeuCysG1ySerValPheLeuValSerGlnLeuPheThr1IleSerProAlaGlnHis 181  
QY 544 GAAGACGGGTGAATGACAGACGAGGAATTGCTGCTGCTTCTTCTTCTATCTTCCTC 603  
Db 182 GluThrValG1yAlaAspCysAsnCysSer1IleTyrProG1yHis1IleThrG1yHisArgMet 201  
QY 604 GCTGGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 627  
Db 202 AlaTrrAspMetMetMetMetMetMetMetMetMetMetMetMetMetMetMetMetMet 209

RESULT 8  
US-08-927-597-48  
Sequence 48, Application US/08927597  
Patent No 6245503  
GENERAL INFORMATION:

Db 102 MetIleMethIstHrProGlyCysValProCysValArgGluAsnSerSerArgCys 121  
QY 364 TGGGTACCGCTCACCCCGCTGAGTACAGAACCCAGCTCCCGCTACGACATA 423  
Db 122 TrpValAlaLeuThrProThrLeuAlaAlaArgAsnAlaSerValProThrThrIle 141  
QY 424 CGAGCCAGCTGCAATTTGCTGCTGGGGCGGCTTCTTCTCCGTATGACGTGGG 483  
Db 142 ArgArgHisValAspLeuValGlyAlaAlaIlePheCysSerAlaMetGlyValGly 161  
QY 484 GATCTGCGGATCTGCTCTCCGCTCCAGCTGTTACACATCTGCGCGCGGCAT 543  
Db 162 AspLeuGlySerValPheLeuValSerGlnLeuPheThrIleSerProArgArgHis 181  
QY 544 GAGACGGTGCAGACACTGCAATTTGCTCAATCATCCGCGCCACATACAGCTACG 603  
Db 182 GluThrValGlnAspCysAsnCysSerIleTyrProGlyHisIleThrGlyHisArgMet 201  
QY 604 GCTTGGATATGATGATGAACCTGC 627  
Db 202 AlaTrpAspMetMetAsnTrp 209

## RESULT 9

US-08-612-973-50  
; Sequence 50, Application US/08612973  
; Patent No. 6150134  
; GENERAL INFORMATION:  
; APPLICANT: MAERTENS, GEERT  
; APPLICANT: BOSMAN, FONS  
; APPLICANT: DE MARTYNOFF, GUY  
; TITLE OF INVENTION: PURIFIED HEPATITIS C VIRUS ENVELOPE  
; TITLE OF INVENTION: PROTEINS FOR DIAGNOSTIC AND THERAPEUTIC USE  
; NUMBER OF SEQUENCES: 111  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: NIXON & VANDERHUYE P.C.  
; STREET: 1100 NORTH GLEBE ROAD  
; CITY: ARLINGTON  
; STATE: VIRGINIA  
; COUNTRY: U.S.A.  
; ZIP: 22201-4714  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/612,973  
; FILING DATE: 11-Mar-1996  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: BYRNE, THOMAS E.  
; REGISTRATION NUMBER: 32,205  
; REFERENCE/DOCKET NUMBER: 1487-10  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (703) 816-4000  
; TELEFAX: (703) 816-4100  
; INFORMATION FOR SEQ ID NO: 50:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 809 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; US-08-612-973-50

Alignment Scores:  
Pred. No.: 1.69e-109  
Score: 1108.00  
Best Local Similarity: 100.00%  
Best Percent Similarity: 99.04%  
Query Match: 95.93%  
DB: 4  
Length: 809  
Matches: 206  
Conservative: 2  
Mismatch: 0  
Indels: 0  
Gaps: 0

## US-09-899-303a-7 (1-633) x US-08-612-973-50 (1-809)

QY 4 TTGGGTAAAGGTATCATATCATACCTTACGTGCGGCTTCCGCGACCTCATGGGTATCTCCG 63  
Db 119 LeuGlyLysValIleAspThrIleThrCysGlyPheAlaAspLeuValGlyTyrIlePro 138  
QY 64 CTCGTGCGCGCCCGCTAGGGGGTGTGCGCAGAGCCCTGGCGCATAGGCGTCCGGTTCTG 123  
Db 139 LeuValGlyAlaProLeuGlyGlyAlaAlaArgAlaLeuAlaHisGlyValArgValLeu 158  
QY 124 GAGACCGCGGTGACACTTGCACAGGGAATTTGGCTGCTGCTTCTTCTATCTTCTC 183  
Db 159 GluAspIleValAsnTyrAlaThrGlyAsnLeuProGlyCysSerPheSerIlePheLeu 178  
QY 184 TTGACCTTACTGTCCTCTGACCATTCACAGCTTCCGCTTATGAGTGGCAGACGTGTC 243  
Db 179 LeuAlaLeuLeuSerCysLeuThrValProAlaSerIleArgGluValArgAsnValSer 198  
QY 244 GGCATGTACCATGTACAGAACGACTGCTCCAACTCAAGCATTTGTATAGAGCGACGAC 303  
Db 199 GlyMetTyrHisValThrAsnAspCysSerAsnSerSerIleValTyrGluAlaAlaAsp 218  
QY 304 ATGATCATGACACACCCCGGGGTGGCTGCTGCTGCTGGGAGAACACTTTCGCTGC 363  
Db 219 MetIleMethIstHrProGlyCysValProCysValArgGluAsnSerSerArgCys 238  
QY 364 TGGGTACCGCTCACCCCGCTGAGTACAGAACCCAGCTCCCGCTACGACATA 423  
Db 239 TrpValAlaLeuThrProThrLeuAlaAlaArgAsnAlaSerValProThrThrIle 258  
QY 424 CGAGCCAGCTGCAATTTGCTGCTGGGGCGGCTTCTTCTCCGTATGACGTGGG 483  
Db 259 ArgArgHisValAspLeuValGlyAlaAlaIlePheCysSerAlaMetGlyValGly 278  
QY 484 GATCTGCGGATCTGCTCTCCGCTCCAGCTGTTACACATCTGCGCGGCAT 543  
Db 279 AspLeuGlySerValPheLeuValSerGlnLeuPheThrIleSerProArgArgHis 298  
QY 544 GAGACGGTGCAGACACTGCAATTTGCTCAATCATCCGCGCCACATACAGCTACG 603  
Db 299 GluThrValGlnAspCysAsnCysSerIleTyrProGlyHisIleThrGlyHisArgMet 318  
QY 604 GCTTGGATATGATGATGAACCTGC 627  
Db 319 AlaTrpAspMetMetAsnTrp 326

## RESULT 10

US-08-927-597-50  
; Sequence 50, Application US/08927597  
; Patent No. 6245503  
; GENERAL INFORMATION:  
; APPLICANT: MAERTENS, GEERT  
; APPLICANT: BOSMAN, FONS  
; APPLICANT: DE MARTYNOFF, GUY  
; TITLE OF INVENTION: PURIFIED HEPATITIS C VIRUS ENVELOPE  
; TITLE OF INVENTION: PROTEINS FOR DIAGNOSTIC AND THERAPEUTIC USE  
; NUMBER OF SEQUENCES: 111  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: NIXON & VANDERHUYE P.C.  
; STREET: 1100 NORTH GLEBE ROAD  
; CITY: ARLINGTON  
; STATE: VIRGINIA  
; COUNTRY: U.S.A.  
; ZIP: 22201-4714  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/927,597  
; FILING DATE:

CLASSIFICATION:  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US 08/612,973  
 FILING DATE: 11-MAR-1996  
 ATTORNEY/AGENT INFORMATION:  
 NAME: BYRNE, THOMAS E.  
 REGISTRATION NUMBER: 32,205  
 REFERENCE/DOCKET NUMBER: 1487-10  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (703) 816-4000  
 TELEFAX: (703) 816-4100  
 INFORMATION FOR SEQ ID NO: 50:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 809 amino acids  
 TYPE: amino acid  
 TOPOLOGY: linear  
 MOLECULE TYPE: protein  
 US-08-927-597-50

Alignment Scores:  
 Pred. No.: 1 69e-109  
 Score: 1108.00  
 Percent Similarity: 100.00%  
 Best Local Similarity: 99.04%  
 Query Match: 95.93%  
 DB: 4  
 Gaps: 0

US-09-899-303A-7 (1-633) x US-08-927-597-50 (1-809)

QY 4 TTGGGTAAAGTATCATGATCCCTTACGTGCGGCTTGCAGCCCTCATGGGGTACATTCG 63  
 DB 119 LeuGlyLysValIleAspHrLeuThrCysGlyPheAlaAspLeuValGlyTrillePro 138  
 QY 64 CTCGTGCGGCCCCCTAGGGGGTCTGCCAGAGCCCTGGCGATGGCGGTCTG 123  
 DB 139 LeuValGlyAlaProLeuValGlyAlaAlaArgAlaLeuAlaHisGlyValAlaArgValLeu 158  
 QY 124 GAAGACGGGGTGAATATGACACAGGGAATTGGCTTGCCTTCTTATCTTCCTC 183  
 DB 159 GluAspGlyValAsnTrpAlaThrGlyAsnLeuProGlyCysSerPheSerIlePheLeu 178  
 QY 184 TTGGCTTACTGTCTGTGACATTCAGATTCGCTTATGAGGTGGCAAGTGTCC 243  
 DB 179 LeuAlaLeuLeuSerCysLeuThrValProAlaSerAlaTrpGluValAlaArgAsnValSer 198  
 QY 244 GGGATGACCATGTGACAGACAGCTGCTCCAACTCAAGCATTTGTATAGAGCGGAC 303  
 DB 199 GlyMetTrpHisValThrAsnAspCysSerAsnSerSerIleValTrpGluAlaAlaAsp 218  
 QY 304 ATGATCATGCACACCCCGGGTGGCGCTTGCAGGAGAACAACTTCCCGCTGC 363  
 DB 219 MetIleMetHisThrProGlyCysValProCysValAlaGlyAsnAsnSerSerArgCys 238  
 QY 364 TTGGTACGCTCACCACCCAGCTCCAGCTAGAGAACGCCAGCGCTCCCACTAGACATA 423  
 DB 239 TrpValAlaLeuThrProThrLeuAlaAlaArgAsnAlaSerValProThrTrpThrIle 258  
 QY 424 CCAAGCCACGCTGATGCTGCTGGGGGGCTCTTGTGTTCCGATGATGAGTGGG 483  
 DB 259 ArgArgHisValAspLeuLeuValGlyAlaAlaAlaPheCysSerAlaMetTrpValGly 278  
 QY 484 GATCTGTGGGATCTGTCTCTCGTCCAGCTTTCACATCTCCGCTGCGCGGAT 543  
 DB 279 AspLeuCysGlySerValPheLeuValSerGlnLeuPheThrIleSerProAlaArgHis 298  
 QY 544 GAGAGCGTCAAGACTGCAATGCTCAATCTATCCGGGCACATAAGAGTACCGTATG 603  
 DB 299 GluThrValGlnAspCysAsnCySerIleTrpProGlyHisIleThrGlnHisAlaGln 318  
 QY 604 GCTTGAGATATGATGATGATGATG 627  
 DB 319 AlaTrpAspMetMetMetAsnTrp 326

RESULT 11  
 US-08-462-195-2

Sequence 2, Application US/08462195  
 Patent No. 5789544

GENERAL INFORMATION:  
 APPLICANT: MIYAMURA, TATSUO  
 APPLICANT: SATO, IZUMI  
 APPLICANT: MATSURA, YOSHIMARU  
 APPLICANT: HONDA, YOSHIKAZU  
 APPLICANT: SEKI, MAKOTO  
 TITLE OF INVENTION: METHOD FOR PRODUCING ECTOPROTEIN OF  
 TITLE OF INVENTION: HEPATITIS C VIRUS  
 NUMBER OF SEQUENCES: 9  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: OBION, SPIVAK, MCCLELLAND, MAIER & NEUSTADT,  
 ADDRESSEE: P.C.  
 STREET: 1755 S. Jefferson Davis Highway, Suite 400  
 CITY: Arlington  
 STATE: Virginia  
 COUNTRY: U.S.A.  
 ZIP: 22202

COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: PatentIn Release #1.0, Version #1.25  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/462,195  
 FILING DATE: 05-JUN-1995  
 CLASSIFICATION: 514  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US/08/446,303  
 FILING DATE: 22-MAY-1995  
 APPLICATION NUMBER: US 08/074,584  
 FILING DATE: 11-JUN-1993  
 APPLICATION NUMBER: JP 152487/1992

ATTORNEY/AGENT INFORMATION:  
 NAME: Obion, No. 5789544man F.  
 REGISTRATION NUMBER: 24,618  
 REFERENCE/DOCKET NUMBER: 4169-003-0  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (703) 413-3000  
 TELEFAX: (703) 413-2220  
 TELEX: 248855 OPAT UR

INFORMATION FOR SEQ ID NO: 2:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 340 amino acids  
 TYPE: amino acid  
 TOPOLOGY: linear  
 MOLECULE TYPE: protein  
 US-08-462-195-2

Alignment Scores:  
 Pred. No.: 4 24e-107  
 Score: 1084.00  
 Percent Similarity: 98.56%  
 Best Local Similarity: 95.67%  
 Query Match: 93.85%  
 DB: 1  
 Gaps: 0

US-09-899-303A-7 (1-633) x US-08-462-195-2 (1-340)

QY 4 TTGGGTAAAGTATCATGATCCCTTACGTGCGGCTTGCAGCCCTCATGGGGTACATTCG 63  
 DB 119 LeuGlyLysValIleAspHrLeuThrCysGlyPheAlaAspLeuValGlyTrillePro 138  
 QY 64 CTCGTGCGGCCCCCTAGGGGGTCTGCCAGAGCCCTGGCGATGGCGGTCTG 123  
 DB 139 LeuValGlyAlaProLeuValGlyAlaAlaArgAlaLeuAlaHisGlyValAlaArgValLeu 158  
 QY 124 GAAGACGGGGTGAATATGACACAGGGAATTGGCTTGCCTTCTTATCTTCCTC 183

Db 159 GlnAspGlyValAsnTyrAlaThrGlyAsnLeuProGlyCysSerPheSerIlePheLeu 178  
QY 184 TTGGCTTACTGTCTGCTGTGCACATTCACAGCTCCGGTTATAGAGTCGCAACGTCTCC 243  
Db 179 LeuAlaLeuLeuSerCysLeuThrIleProAlaSerAlaTyrGluValArgAsnValSer 198  
QY 244 GGGATGTACACATGTACAGACAGACATGCTCCAACTCAACATTTGTGTATAGAGCAGCGAC 303  
Db 199 GlyIleTyrHisValThrAsnAspCysSerAsnSerIleValTyrGluAlaAlaAsp 218  
QY 304 ATGATCATGACACACCCCGGGGTGCGCTCCGCTTCGGAGAACACATCTCCCGCGTGC 363  
Db 219 ValIleMetHisAlaProGlyCysValProCysValArgGluAsnAsnSerIleArgCys 238  
QY 364 TGGGTAGCGCTACACCCCGACGCTGCAGCTAGAAAGCCAGCGTCCGACATACGACAAATA 423  
Db 239 TrpValAlaLeuThrProThrLeuAlaAlaArgAsnAlaSerValProThrThrIleu 258  
QY 424 CGAGCCGACGTCGATTCCTGCTGGGGCGGCTGCTTCTGTCCTATGACGTGGGG 483  
Db 259 ArgArgHisValAspLeuValGlyThrAlaAlaPheCysSerAlaMetTyrValGly 278  
QY 484 GATCTCTCGGATCTGTCTCTCTCCAGCTGCACATCTCCGCTCCGCGGAT 543  
Db 279 AspLeuCysGlySerValPheLeuIleSerGlnLeuPheThrPheSerProArgArgHis 298  
QY 544 GAGACGGTGCAGAGACTGCAATTGCTCAATATATCCGGCCACATACAGTACCGCTATG 603  
Db 299 GluThrValGlnAspCysAsnCysSerIleTyrProGlyHisValSerGlyHisArgMet 318  
QY 604 GCTTGGGATATGATGATGAACACTGG 627  
Db 319 AlaTrpAspMetMetMetAsnTrp 326  
RESULT 12  
US-08-636-883-2  
Sequence 2, Application US/08636883  
Patent No. 5830691  
GENERAL INFORMATION:  
APPLICANT: MIYAMURA, TATSUO  
APPLICANT: SAITO, IZUMU  
APPLICANT: MATSUURA, YOSHIMARU  
APPLICANT: HONDA, YOSHIKAZU  
APPLICANT: SEKI, MAKOTO  
TITLE OF INVENTION: METHOD FOR PRODUCING ECTOPROTEIN OF  
NUMBER OF SEQUENCES: 9  
CORRESPONDENCE ADDRESSES:  
ADDRESSEE: OBLON, SPIVAK, MCCLELLAND, MAIER & NEUSTADT,  
ADDRESSEE: P.C.  
STREET: 1755 S. Jefferson Davis Highway, Suite 400  
CITY: Arlington  
STATE: Virginia  
COUNTRY: U.S.A.  
ZIP: 22202  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/636,883  
FILING DATE:  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US/08/446,303  
FILING DATE:  
APPLICATION NUMBER: US 08/074,584  
FILING DATE: 11-JUN-1993  
APPLICATION NUMBER: JP 152487/1992  
FILING DATE: 11-JUN-1992  
ATTORNEY/AGENT INFORMATION:  
NAME: Oblon, No. 5830691man F.

REGISTRATION NUMBER: 24,618  
REFERENCE/DOCKET NUMBER: 4169-003-0  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (703) 413-3000  
TELEFAX: (703) 413-2220  
TELEX: 248855 OPAT UR  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 340 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-636-883-2  
Alignment Scores:  
Pred. No.: 4,24e-107 Length: 340  
Score: 1084.00 Matches: 199  
Percent Similarity: 98.56% Conservative: 6  
Best Local Similarity: 95.67% Mismatches: 3  
Query Match: 93.85% Indels: 0  
DB: 2 Gaps: 0  
US-09-899-303A-7 (1-633) x US-08-636-883-2 (1-340)  
QY 4 TTGGGTAAAGTCAATGATACCTTACGTGCGGCTTCCGCGCACTTCATGGGTACATTCCG 63  
Db 119 LeuGlyValAlaAspThrLeuThrCysGlyPheAlaAspLeuMetGlyTyrIlePro 138  
QY 64 CTCGTCGCGGCGCCCGCTAGGGGGTCTCCAGAGCCCTGGGCGCATAGGGCTCGGGTTGTG 123  
Db 139 LeuValGlyAlaProLeuGlyGlyAlaAlaArgAlaLeuAlaHisGlyValArgValLeu 158  
QY 124 GAAGACGGCGTGAACCTATGCAACAGAGGAATTTCGCTGGCTCTTCTATGTCCTC 183  
Db 159 GlnAspGlyValAsnTyrAlaThrGlyAsnLeuProGlyCysSerPheSerIlePheLeu 178  
QY 184 TTGGCTTACTGTCTGCTGTGCACATTCACAGCTCCGGTTATAGAGTCGCAACGTCTCC 243  
Db 179 LeuAlaLeuLeuSerCysLeuThrIleProAlaSerAlaTyrGluValArgAsnValSer 198  
QY 244 GGGATGTACACATGTACAGACAGACTGCTCCAACTCAACATTTGTGTATAGAGCAGCGAC 303  
Db 199 GlyIleTyrHisValThrAsnAspCysSerAsnSerIleValTyrGluAlaAlaAsp 218  
QY 304 ATGATCATGACACACCCCGGGGTGCGCTCCGCTTCGGAGAACACATCTCCCGCGTGC 363  
Db 219 ValIleMetHisAlaProGlyCysValProCysValArgGluAsnAsnSerIleArgCys 238  
QY 364 TGGGTAGCGCTACACCCCGACGCTGCAGCTAGAAAGCCAGCGTCCGACATACGACAAATA 423  
Db 239 TrpValAlaLeuThrProThrLeuAlaAlaArgAsnAlaSerValProThrThrIleu 258  
QY 424 CGAGCCGACGTCGATTCCTGCTGGGGCGGCTGCTTCTGTCCTATGACGTGGGG 483  
Db 259 ArgArgHisValAspLeuValGlyThrAlaAlaPheCysSerAlaMetTyrValGly 278  
QY 484 GATCTCTCGGATCTGTCTCTCTCCAGCTGCACATCTCCGCTCCGCGGAT 543  
Db 279 AspLeuCysGlySerValPheLeuIleSerGlnLeuPheThrPheSerProArgArgHis 298  
QY 544 GAGACGGTGCAGAGACTGCAATTGCTCAATATATCCGGCCACATACAGTACCGCTATG 603  
Db 299 GluThrValGlnAspCysAsnCysSerIleTyrProGlyHisValSerGlyHisArgMet 318  
QY 604 GCTTGGGATATGATGATGAACACTGG 627  
Db 319 AlaTrpAspMetMetMetAsnTrp 326  
RESULT 13  
US-09-127-829-2  
Sequence 2, Application US/09127829  
Patent No. 6063904  
GENERAL INFORMATION:

```

APPLICANT: MIYAMURA, TATSUO
APPLICANT: SAITO, IZUMU
APPLICANT: MATSURA, YOSHIMARU
APPLICANT: HONDA, YOSHITAKU
APPLICANT: SEKI, MAKOTO
TITLE OF INVENTION: METHOD FOR PRODUCING ECTOPROTEIN OF
TITLE OF INVENTION: HEPATITIS C VIRUS
NUMBER OF SEQUENCES: 9
CORRESPONDENCE ADDRESS:
ADDRESSEE: OBLON, SPIVAK, MCLELLAND, MAIER & NEUSTADT,
ADDRESSEE: P.C.
STREET: 1755 S. Jefferson Davis Highway, Suite 400
City: Arlington
STATE: Virginia
COUNTRY: U.S.A.
ZIP: 22202
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/127,829
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/446,303
FILING DATE:
APPLICATION NUMBER: US 08/074,584
FILING DATE: 11-JUN-1993
APPLICATION NUMBER: JP 152487/1992
FILING DATE: 11-JUN-1992
ATTORNEY/AGENT INFORMATION:
NAME: Oblon, No. 6063904man F.
REGISTRATION NUMBER: 24,618
REFERENCE/DOCKET NUMBER: 4169-003-0
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 413-3000
TELEFAX: (703) 413-2220
TELEX: 248855 OPAT UR
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 340 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-09-127-829-2

Alignment Scores:
Pred. NO.: 4.24e-107 Length: 340
Score: 1084.00 Matches: 199
Percent Similarity: 98.56% Conservative: 6
Best Local Similarity: 95.67% Mismatches: 3
Query Match: 93.85% Indels: 0
DB: 3 Gaps: 0

US-09-899-303a-7 (1-633) x US-09-127-829-2 (1-340)
QY 4 TTGGGTAAGCATCATGATACCTTACCTTACGCGGCTTGCGCGACCTCATGGGGTACATTCCG 63
Db 119 LeucllylgsvalilleaprlhleprthleuthCys6lyPhehlaalspuewetgLyTylllepio 138
QY 64 CTCGTGCGCGCCCCCTAGGGGGGTGCTGCAGAGCCCTGGCGGATGGCGCTCCGCGTCTTG 1233
Db 139 LeuValGlyAlaProlaueGlyGlyAlaAlaArgAlaLeuAlaIhlsGlyValAlaArgValLeu 158
QY 124 GAACAGCGCGTGAACCTATGCACACGGGAATTTGGCTGCTGCTCTTTCTATCTTCCTC 1833
Db 159 GluAspGlyValAlaenTyAlaIhthGlyAsnLeuProGlyCysSerPheSerIlePheLeu 178
QY 184 TTGGCTTACGTCCTGCTGACCATTCAGCTTCGCGCTTATGAGGTGGCGGCAACGCTGC 243
Db 179 LeuAlaLeuLeuSerCysLeuThIleProlaSerAlaTyGlyValAlaArgAsnValSer 198

```

OY	244	GGGATAGACCATGTGCAGAACGACTGGTCCAACTCAAGATTGTGATAGGACGACG	303
Db	199	GlytelyrhstvaltnrshnabpcyssehnserSettlevalytlUnlaIalasp	218
OY	304	ATGATCATGCACACCSCCGGGTGCGTGTGCCGTTCGGGAGAACAACTTTCCCGTGC	363
Db	219	ValllEmethIsAlaProgljCysValProCysValArgGlunashnSerterArgcys	238
OY	364	TGGTAGCGGCTCACCCCCACGCTCGCAGCTAGGAACGCCAGGTCGCCACTACGAAATA	423
Db	239	tTrpAlaLeuThrPrroThlrleuaIalaIargSnAlaSeValPrrThrThrleu	258
OY	424	CGAGCGCCACGTCGATTTGTCGTCGTGGGGCGGCGTCCTTGTTCGGCTATGACSTGGG	483
Db	259	ArgArglnshValasprleuleuValgUthrAlaAlaPhncysSerAlametyltValgU	278
OY	484	GATCTGTCCGAGATCTGTCTTCCTGTCTCCATCCAGCTTTGACCATCTCCGCTCCCGGAT	543
Db	279	AsprEncycsglySerValPheleulleSerGlnLeuPheThrPheserProArglnhs	298
OY	544	GAGACGGTGCAGACGCTGCATTCGTAATCTACCCGGCCACAATAACAGTACCCTATG	603
Db	299	GIuthrvAlgnlsnpCysasnCyssertletyrProgljnlshValserclYhlslrgmet	318
OY	604	GCTTGGGATATGATGATGACTGG 627	
Db	319	Alatrpsarmetmetmetantrp 326	
 RESULT 14 US-09-014-416-3 Sequence 3, Application US/09014416 Patent No. 6153421 GENERAL INFORMATION: APPLICANT: Yanagil, Masayuki APPLICANT: Bukh, Jems APPLICANT: Emerson, Susanne U. TITLE OF INVENTION: CLODED GENOMES OF INFECTIOUS HEPATITIS C VIRUSES AND TITLE OF INVENTION: USES THEREOF FILE REFERENCE: 20264276 CURRENT APPLICATION NUMBER: US/09/014,416 CURRENT FILING DATE: 1998-01-27 EARLIER APPLICATION NUMBER: US 60/053,062 EARLIER FILING DATE: 1997-07-18 NUMBER OF SEQ ID NOS: 65 SOFTWARE: PatentIn Ver. 2.1 SEQ ID NO 3 LENGTH: 3010 TYPE: PRT ORGANISM: Hepatitis C virus US-09-014-416-3			
 Alignment Scores:			
Pred. NO.:	2,17e-106	Length:	3010
Score:	1081.00	Matches:	199
Percent Similarity:	98.56%	Conservative:	6
Best Local Similarity:	95.67%	Mismatches:	3
Query Match:	93.59%	Indels:	0
DB:	4	Gaps:	0
 US-09-899-303A-7 (1-633) x US-09-014-416-3 (1-3010)			
OY	4	TTGGGTAGAAGTCATCATGATACCCCTTACGTGCGGGCTTGGCCAGCTCATGGGGATATTCCG	63
Db	119	leucdilyustValilleasprhtlleuthrCysglYphneAlaAsprleumec[s]ytUtllepro	138
OY	64	CTGCTGCGGCCCCCCTTAGGGGGGTCTGTCGCAGAGCCCTTGCGCATAGSGCTCCGGGTTTG	123
Db	139	leuValaIgluAlaroleuglscylalaaIaladgrAlaleuAlahlglslyalalargyaldeu	158
OY	124	GAAACAGCGGCTGAACATATGCAADAAGGAAATTGCGGTGGTGGTCTTTGCTATCTTCTGC	183

```

Db 159 GluaspclYalaAnTYrAlaThrGlyAsnLeuProGlyCysSerPheSerIlePheLeu 178
QY 184 TTGGCTTTACTGTCCTGTCTGACCAATTCACGCTTCGCTTATGAGCGGCAACGTCCTC 243
Db 179 LeuAlaLeuLeuSerCysLeuThrIleProAlaSerAlaTYrGluValAlaGsnValSer 198
QY 244 GGGATGTACCATGTGCAGAGAGACTGCTCCAACTCAAGCATTTGTGTATGAGGCGGCGAC 303
Db 199 GlyIleTYrHisValIThrAsnAspCysSerAsnSerIleValTYrGluAlaAlaAsp 218
QY 304 ATGATCATMGACACACCCCGGGTGGCGCTCCGCTTCGGGAGAACACTCTTCCCGCTGC 363
Db 219 ValIleMetHisThrProGlyCysValProCysValGlnGluGlyAsnSerSerAlaGys 238
QY 364 TGGGTACCGCTCACCCCGCTGCGAGCTAGAGAACGCCAGCGTCCCACTACGACATA 423
Db 239 TrpValAlaLeuThrProThrIleuAlaAlaArgsnAlaSerValProThrThrIle 258
QY 424 CGAGCGCACGTCGATTTGCTGCTGGGGCGGCTCTTCTGTTCGCTATGATACGTGGG 483
Db 259 ArgArgHisValAspLeuLeuValGlyThrAlaAlaPheCysSerAlaMetTYrValGly 278
QY 484 GATCTCGCGGATGTCCTTCCTGCTCCAGCTGTTCACCATCTGCGCTCGCGGCGAT 543
Db 279 AspLeuCysGlySerIlePheLeuValSerGlnLeuPheThrPheSerProAlaArgHis 298
QY 544 GAGACGGTGCAGACATTCGCTCAATCTATCCGCGCCACATACAGAGTCAACGCTATG 603
Db 299 GluThrValGlnAspCysAsnCysSerIleTYrProGlyHisValSerGlyHisArgMet 318
QY 604 GCTTGGATATGATGATGAACCTGG 627
Db 319 AlaTrpAspMetMetAlaSnTrp 326

```

## RESULT 15

```

US-09-014-416-5
: Sequence 5, Application US/09014416
: Patent No. 6153421
: GENERAL INFORMATION:
: APPLICANT: Yanagi, Masayuki
: APPLICANT: Bukh, Jens
: APPLICANT: Emerson, Susanne U.
: APPLICANT: Purcell, Robert H.
: TITLE OF INVENTION: CLONED GENOMES OF INFECTIOUS HEPATITIS C VIRUSES AND
: FILE REFERENCE: 20264276
: CURRENT APPLICATION NUMBER: US/09/014,416
: EARLIER FILING DATE: 1998-01-27
: EARLIER APPLICATION NUMBER: US 60/053,062
: NUMBER OF SEQ ID NOS: 65
: SOFTWARE: PatentIn Ver. 2.1
: SEQ ID NO 5
: LENGTH: 3011
: TYPE: PRT
: ORGANISM: Hepatitis C virus
US-09-014-416-5

```

## Alignment Scores:

```

Pred. No.: 2.17e-106 Length: 3011
Score: 1081.00 Matches: 199
Percent Similarity: 98.56% Conservative: 6
Best Local Similarity: 95.67% Mismatches: 3
Query Match: 93.59% Indels: 0
DB: 4 Gaps: 0

```

US-09-899-303a-7 (1-633) x US-09-014-416-5 (1-3011)

```

QY 4 TTGGGTAAAGTCATGATACCTTACCTTACGCGCTTCGCCGACATCAATGGGCTACATTCG 63
Db 119 LeuclYysValIleAspThrLeuThrCysGlyPheAlaAspLeuMetGlyTYrIlePro 138
QY 64 CTCGTGCGCGCCCCCTAGGGGGGTGTGCCAGAGCCCTGGCGCATGGCGGTCCGGGTTCTG 123

```

```

Db 139 LeuValAlaLaProLeuGlyGlyAlaAlaArgAlaLeuAlaHisGlyValAlaValLeu 158
QY 124 GAAGACGCGCGTGAACATATGCAACAGGAATTTGGCTGGTCTCTTCTATCTTCCTC 183
Db 159 GluaspclYalaAnTYrAlaThrGlyAsnLeuProGlyCysSerPheSerIlePheLeu 178
QY 184 TTGGCTTTACTGTCCTGTCTGACCAATTCACGCTTCGCTTATGAGCGGCAACGTCCTC 243
Db 179 LeuAlaLeuLeuSerCysLeuThrIleProAlaSerAlaTYrGluValAlaGsnValSer 198
QY 244 GGGATGTACCATGTGCAGAGAGACTGCTCCAACTCAAGCATTTGTGTATGAGGCGGCGAC 303
Db 199 GlyIleTYrHisValIThrAsnAspCysSerAsnSerIleValTYrGluAlaAlaAsp 218
QY 304 ATGATCATMGACACACCCCGGGTGGCGCTCCGCTTCGGGAGAACACTCTTCCCGCTGC 363
Db 219 ValIleMetHisThrProGlyCysValProCysValGlnGluGlyAsnSerSerAlaGys 238
QY 364 TGGGTACCGCTCACCCCGCTGCGAGCTAGAGAACGCCAGCGTCCCACTACGACATA 423
Db 239 TrpValAlaLeuThrProThrIleuAlaAlaArgsnAlaSerValProThrThrIle 258
QY 424 CGAGCGCACGTCGATTTGCTGCTGGGGCGGCTCTTCTGTTCGCTATGATACGTGGG 483
Db 259 ArgArgHisValAspLeuLeuValGlyThrAlaAlaPheCysSerAlaMetTYrValGly 278
QY 484 GATCTCGCGGATGTCCTTCCTGCTCCAGCTGTTCACCATCTGCGCTCGCGGCGAT 543
Db 279 AspLeuCysGlySerIlePheLeuValSerGlnLeuPheThrPheSerProAlaArgHis 298
QY 544 GAGACGGTGCAGACATTCGCTCAATCTATCCGCGCCACATACAGAGTCAACGCTATG 603
Db 299 GluThrValGlnAspCysAsnCysSerIleTYrProGlyHisValSerGlyHisArgMet 318
QY 604 GCTTGGATATGATGATGAACCTGG 627
Db 319 AlaTrpAspMetMetAlaSnTrp 326

```

Search completed: February 12, 2003, 08:56:18  
Job time : 21 secs

**THIS PAGE BLANK (USPTO)**



PT produce proteins suitable for direct use in vaccines or diagnostic assays of HCV

XX Claim 23; Fig 21; 146pp; English.

XX PS  
XX AAT12704-T12709 and AAT12961-T12974 represent hepatitis C virus (HCV) E1 and E2 protein coding sequence constructs. These sequences are included in vectors for the production of recombinant E1, E2, and E1/E2 proteins. CC The recombinant proteins can then be isolated using a method of the invention. In the method, the envelope proteins are purified by carrying out a disulphide bond cleavage, or a reduction step with a disulphide bond cleavage agent, after lysis of recombinant host cells. CC The constructs containing the purified HCV envelope proteins can be used for vaccinating humans against HCV, for in vitro detection of HCV antibodies in a sample, and in a serotyping assay for detecting one or more serological types of HCV present in a biological sample. The CC constructs can also be immobilised on a solid substrate and incorporated into a reversed phase hybridisation assay for determining the presence or the genotype of HCV. The new purification method preserves the CC conformation of the recombinantly expressed E1, E2 and E1/E2, and CC eliminates contaminating proteins. Antigens isolated using this method CC are more reactive with human sera than those isolated by known CC techniques.

SQ Sequence 633 BP; 111 A; 192 C; 174 G; 156 T; 0 other;

Query Match 100.0%; Score 633; DB 17; Length 633;

Best Local Similarity 100.0%; Pred. No. 5,7e-166;

Matches 633; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

OY 1 ATGTGGGTAAAGTATGATACCTTACGTCGCGCTTCGCCGACCTATGGGGTACTT 60
DB 1 ATGTGGGTAAAGTATGATACCTTACGTCGCGCTTCGCCGACCTATGGGGTACTT 60
OY 61 CCGCTCGTCGGCGCCCGCTAGGGGCTCTGCAGAGCCCTGCGCCATGAGCGCTCCGGGTT 120
DB 61 CCGCTCGTCGGCGCCCGCTAGGGGCTCTGCAGAGCCCTGCGCCATGAGCGCTCCGGGTT 120
OY 61 CCGCTCGTCGGCGCCCGCTAGGGGCTCTGCAGAGCCCTGCGCCATGAGCGCTCCGGGTT 120
DB 61 CCGCTCGTCGGCGCCCGCTAGGGGCTCTGCAGAGCCCTGCGCCATGAGCGCTCCGGGTT 120
OY 121 CTGGAAGACGGCGTGAATATGCAACAGGAAATTTGCTGTTCTCTATCTTC 180
DB 121 CTGGAAGACGGCGTGAATATGCAACAGGAAATTTGCTGTTCTCTATCTTC 180
OY 121 CTGGAAGACGGCGTGAATATGCAACAGGAAATTTGCTGTTCTCTATCTTC 180
DB 121 CTGGAAGACGGCGTGAATATGCAACAGGAAATTTGCTGTTCTCTATCTTC 180
OY 181 CTCTGGTTCATCTCTCTCTGTCGTCATTCAGCTTCGCTTATGAGTGGCAAGCTG 240
DB 181 CTCTGGTTCATCTCTCTCTGTCGTCATTCAGCTTCGCTTATGAGTGGCAAGCTG 240
OY 181 CTCTGGTTCATCTCTCTCTGTCGTCATTCAGCTTCGCTTATGAGTGGCAAGCTG 240
DB 181 CTCTGGTTCATCTCTCTCTGTCGTCATTCAGCTTCGCTTATGAGTGGCAAGCTG 240
OY 241 TCCGGAGTACATGATGCAACGACACGCTGTCACACTCAAGATGTGTATGAGCAGCG 300
DB 241 TCCGGAGTACATGATGCAACGACACGCTGTCACACTCAAGATGTGTATGAGCAGCG 300
OY 241 TCCGGAGTACATGATGCAACGACACGCTGTCACACTCAAGATGTGTATGAGCAGCG 300
DB 241 TCCGGAGTACATGATGCAACGACACGCTGTCACACTCAAGATGTGTATGAGCAGCG 300
OY 301 GACATGATCATGACACACCCCGGGTGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 360
DB 301 GACATGATCATGACACACCCCGGGTGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 360
OY 301 GACATGATCATGACACACCCCGGGTGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 360
DB 301 GACATGATCATGACACACCCCGGGTGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 360
OY 361 TGTCTGGTGAAGGCTACACCCCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 420
DB 361 TGTCTGGTGAAGGCTACACCCCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 420
OY 361 TGTCTGGTGAAGGCTACACCCCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 420
DB 361 TGTCTGGTGAAGGCTACACCCCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 420
OY 421 ATACACACCCAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 480
DB 421 ATACACACCCAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 480
OY 421 ATACACACCCAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 480
DB 421 ATACACACCCAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 480
OY 481 GGGGATCTCTGGGATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 540
DB 481 GGGGATCTCTGGGATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 540
OY 481 GGGGATCTCTGGGATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 540
DB 481 GGGGATCTCTGGGATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 540
OY 541 CATGAGAGGTCAGAGTATGCAATGCTCAATCTATCCCGGCAATCAAGTACAGTCCGCT 600
DB 541 CATGAGAGGTCAGAGTATGCAATGCTCAATCTATCCCGGCAATCAAGTACAGTCCGCT 600
OY 541 CATGAGAGGTCAGAGTATGCAATGCTCAATCTATCCCGGCAATCAAGTACAGTCCGCT 600
DB 541 CATGAGAGGTCAGAGTATGCAATGCTCAATCTATCCCGGCAATCAAGTACAGTCCGCT 600
OY 601 ATGGCTTGGGATATGATGATGAATGTAATAG 633
DB 601 ATGGCTTGGGATATGATGATGAATGTAATAG 633
OY 601 ATGGCTTGGGATATGATGATGAATGTAATAG 633
DB 601 ATGGCTTGGGATATGATGATGAATGTAATAG 633

```

# RESULT 2

AAT12709

ID AAT12709 standard; DNA; 636 BP.

XX AAT12709;

XX 23-SEP-1996 (first entry)

DE HCV E1 construct HCC117A.

XX HCV; E1, E2; disulphide bond cleavage; envelope protein; vaccine; human; serotype; reversed phase hybridisation assay; genotype; antigen; sera;

XX ss.

OS Hepatitis C virus.

PN W09604385-A2.

XX 15-FEB-1996.

PP 31-JUL-1995; 95WO-EP03031.

PR 29-JUL-1994; 94EP-0870132.

XX (INNO-) INNOGENETICS NV.

PI Bosman F, Buyse M, De Martynoff G, Maertens G;

XX WPI: 1996-129401/13.

PT purifying recombinant hepatitis C virus (HCV) E1 and E2 envelope proteins - in presence of disulphide bond cleavage agent, to produce proteins suitable for direct use in vaccines or diagnostic assays of HCV

PS Claim 23; Fig 21; 146pp; English.

XX AAT12704-T12709 and AAT12961-T12974 represent hepatitis C virus (HCV) E1 and E2 protein coding sequence constructs. These sequences are included in vectors for the production of recombinant E1, E2, and E1/E2 proteins. CC The recombinant proteins can then be isolated using a method of the invention. In the method, the envelope proteins are purified by carrying out a disulphide bond cleavage, or a reduction step with a disulphide bond cleavage agent, after lysis of recombinant host cells. CC The constructs containing the purified HCV envelope proteins can be used for vaccinating humans against HCV, for in vitro detection of HCV antibodies in a sample, and in a serotyping assay for detecting one or more serological types of HCV present in a biological sample. The CC constructs can also be immobilised on a solid substrate and incorporated into a reversed phase hybridisation assay for determining the presence or the genotype of HCV. The new purification method preserves the CC conformation of the recombinantly expressed E1, E2 and E1/E2, and CC eliminates contaminating proteins. Antigens isolated using this method CC are more reactive with human sera than those isolated by known CC techniques.

SQ Sequence 636 BP; 111 A; 197 C; 175 G; 153 T; 0 other;

Query Match 96.8%; Score 612.8; DB 17; Length 636;

Best Local Similarity 96.1%; Pred. No. 2.3e-160;

Matches 620; Conservative 0; Mismatches 12; Indels 0; Gaps 0;

```

OY 1 ATGTGGGTAAAGTATGATACCTTACGTCGCGCTTCGCCGACCTATGGGGTACTT 60
DB 1 ATGTGGGTAAAGTATGATACCTTACGTCGCGCTTCGCCGACCTATGGGGTACTT 60
OY 61 CCGCTCGTCGGCGCCCGCTAGGGGCTCTGCAGAGCCCTGCGCCATGAGCGCTCCGGGTT 120
DB 61 CCGCTCGTCGGCGCCCGCTAGGGGCTCTGCAGAGCCCTGCGCCATGAGCGCTCCGGGTT 120
OY 61 CCGCTCGTCGGCGCCCGCTAGGGGCTCTGCAGAGCCCTGCGCCATGAGCGCTCCGGGTT 120
DB 61 CCGCTCGTCGGCGCCCGCTAGGGGCTCTGCAGAGCCCTGCGCCATGAGCGCTCCGGGTT 120
OY 121 CTGGAAGACGGCGTGAATATGCAACAGGAAATTTGCTGTTCTCTATCTTC 180
DB 121 CTGGAAGACGGCGTGAATATGCAACAGGAAATTTGCTGTTCTCTATCTTC 180

```

Db	121	CTGGAAAGACGGCGTGAACATATGCAAACAGGGAATTGGCTGGTGCCTTTCTCTATCTTC	180
Qy	181	CTCTTTGGCTTACTCTGTCTGTCTTACCATTTCACGCTTCGGCTTATGAGGTGGCGAACGTG	240
Db	181	CTCTTGGCTTACTAGTCTGTGCTTAAACATTTCACGCTTCGGCTTACGAGGTGGCGAACGTG	240
Qy	241	TCCGGAGTATCAATGTCACGAACGACGTGGCCAACTCAAGCAANTGTGATGAGGACGCG	300
Db	241	TCCGGAGTATCAATGTCACGAACGACGTGGCCAACTCAAGCAANTGTGATGAGGACGCG	300
Qy	301	GACATGATCATGCACACCCCGGGGTGGGTGGCTTCGCTTGGGGAGAACATCTTCCCGC	360
Db	301	GACATGATCATGCACACCCCGGGGTGGGTGGCTTCGCTTGGGGAGAACATCTTCCCGC	360
Qy	361	TGCTGGGTAGCGCTCACCCCCAGCGCTGCAGACTAGGAAGCCACGCTCCCCACTACGACA	420
Db	361	TGCTGGGTAGCGCTCACCCCCAGCGCTGCAGACTAGGAAGCCACGATCCCCACTACACACA	420
Qy	421	ATACGAGCGCCACGTCGATTTGCTCTGTGGGGGGGCGCTTTCGATTCGGCTATGTAAGCTG	480
Db	421	ATACGAGCGCCACGTCGATTTGCTCTGTGGGGGGGCGCTTTCGATTCGGCTATGTAAGCTG	480
Qy	481	GGGGATCTCTCGGATCTGTTCTCTGTCTCCAGCTGTTCACACATCTCGCTCGGCGG	540
Db	481	GGGGATCTCTCGGATCTGTTCTCTGTCTCCAGCTGTTCACACATCTCGCTCGGCGG	540
Qy	541	CATGACGAGTGCAGAGACATGCAATTGGTCAATCTATTCGCCGGCCACATTAACAGGTCAACGT	600
Db	541	CATGAGACGGTGCAGAGATGCAATTGGTCAATCTATTCGCCGGCCACATTAACAGGTCAACGT	600
Qy	601	ATGGCTTGGGATGATGATGATGAATCACTGGTAATA 632	
Db	601	ATGGCTTGGGATGATGATGATGAATCACTGGTACTATA 632	

XX	RESULT 3
XX	AA112705
ID	AA112705 standard; DNA; 795 BP.
XX	
XX	AA112705;
AC	
DT	23-SEP-1996 (first entry)
XX	
DE	HCV E1 construct HCC110A.
XX	
KW	HCV; E1; E2; disulphide bond cleavage; envelope protein; vaccine; human
KW	serotype; reversed phase hybridisation assay; genotype; antigen; sera;
KW	ss.
XX	
OS	Hepatitis C virus.
XX	
PN	W09604385-A2.
XX	
PD	15-FEB-1996.
XX	
PE	31-JUL-1995; 95MO-EP03031.
XX	
PR	29-JUL-1994; 94EP-0870132.
XX	
PA	(INNO-) INNOGENETICS NV.
XX	
XX	
PI	Bosman F, Buyse M, De Martynoff G, Maertens G;
XX	
DR	WPI; 1996-129401/13.
XX	
PT	Purifying recombinant hepatitis C virus (HCV) E1 and E2 envelope
PT	proteins - in presence of di: sulphide bond cleavage agent, to
PT	produce proteins suitable for direct use in vaccines or diagnostic
PT	assay of HCV
XX	
PS	Claim 23; Fig 21; 146pp; English.
XX	

CC AA1127/04-T127/09 and AA112961-T12974 represent hepatitis C virus (HCV) E1  
CC and E2 protein coding sequence constructs. These sequences are included  
CC in vectors for the production of recombinant E1, E2, and E1/E2 proteins.  
CC The recombinant proteins can then be isolated using a method of the  
CC invention. In the method, the envelope proteins are purified by  
CC carrying out a disulphide bond cleavage, or a reduction step with a  
CC disulphide bond cleavage agent, after lysis of recombinant host cells.  
CC The constructs containing the purified HCV envelope proteins can be used  
CC for vaccinating humans against HCV, for *in vitro* detection of HCV  
CC antibodies in a sample, and in a serotyping assay for detecting one or  
CC more serological types of HCV present in a biological sample. The  
CC constructs can also be immobilised on a solid substrate and incorporated  
CC into a reversed phase hybridisation assay for determining the presence of  
CC the genotype of HCV. The new purification method preserves the  
CC conformation of the recombinantly expressed E1, E2 and E1/E2, and  
CC eliminates contaminating proteins. Antigens isolated using this method  
CC are more reactive with human sera than those isolated by known  
CC techniques.

50 Sequence 795 BP; 130 A; 240 C; 231 G; 194 T; 0 other;

Query Match	Score	DB	Length
96.28;	608.8;	17;	795;

Matches 616; Conservative 0; Mismatches 12; Indels 0; Gaps 0;

OY	1	ATGTTGGGTAAGGTCATCGATGAACCTTACGAGGCGGCTTCGCCAGCCATGAGGGTACATT	60
Db	1	ATGTTGGGTAAGGTCATCGATGAACCTTACATCGCGCTTCCGCCACCTCGTGGGGTACATT	60
OY	61	CCGCTCGTGCGGCGCCCCCTAAGGGGGTCTGCCAGAGCCCTGGCGCATGGCGCTCCGGGTT	120
Db	61	CCGCTCGTGCGGCGCCCCCTAAGGGGGGCTGCCAGAGGCCCTGGCGCATGGCGTCCGGGTT	120
OY	121	CTGGAAGAGCGGCGTGAATGACAAGAGGAATTGGCTGGTGGCTTTCTCTATCTTC	180
Db	121	CTGGAAGAGCGGCGTGAATGACAAGAGGAATTGGCGGTGGCTTTCTCTATCTTC	180
OY	181	CTCTTGCGTTACTGCTCCTGCTGACACATTCCAGCTTCCGCTTATGAGTGCGCCACGTG	240
Db	181	CTCTTGCGTTGCTGCTCCTGCTGACACGTTCCAGCTTCCGCTTATGAGTGCGCCACGTG	240
OY	241	TCCGGGATGTACATGTTCAGAACAGACTGCTCCAACTCAAGCATTTGTATGAGGACGC	300
Db	241	TCCGGGATGTACCATGTTCACAAAGAGACTGCTCCAACTCAAGCATTTGTATGAGGACGC	300
OY	301	GACATGATCATGCAACACCCCGGGGTGCGCTGCGCTTGGGGAGAACAACCTTCCCGC	360
Db	301	GACATGATCATGCAACACCCCGGGGTGGGTGGCCGTGGGGGAACAACCTTCCCGC	360
OY	361	TGCTGGGATGAGGCTCACCCCCAGCCTGCGACGTGAAAGCCAGACCGCCCACTACGACA	420
Db	361	TGCTGGGATGAGGCTCACCCCCAGCCTGCGACGTGAAAGCCAGACCGTCCCACTACGACA	420
OY	421	ATAGCAGCCCAAGTCGATTTGCTGTTGGGGCGGCTGCTTCTGTTCCGCTATGTAAGTG	480
Db	421	ATAGCAGCCCAAGTCGATTTGCTGTTGGGGCGGCTGCTTCTGTTCCGCTATGTAAGTG	480
OY	481	GGGAGATCTCTCGGATCTGTTTCTCTGCTGCCAGCTGTTACACATCTGCGCTCGCGG	540
Db	481	GGGAGACCTCTCGGATCTGTTTCTCTGCTGCCAGCTGTTTACACATCTGCGCTCGCGG	540
OY	541	CATGAGACGGTGAAGGAGCTGCAATTTGTCATATCCGGGCCACATTAACAGGTCACCG	600
Db	541	CATGAGACGGTGAAGGAGCTGCAATTTGTCATATCTATCCGGGCCACATTAACAGGTCACCG	600
OY	601	ATGGCTTGGGATATGATGATGAACCTGGT	628
Db	601	ATGGCTTGGGATATGATGATGAACCTGGT	628
RESULT 4			
AAAT12973			
AAAT12973 standard; DNA: 2086 BP			



CC carrying out a disulphide bond cleavage, or a reduction step with a  
CC disulphide bond cleavage agent, after lysis of recombinant host cells.  
CC The constructs containing the purified HCV envelope proteins can be used  
CC for vaccinating humans against HCV, for in vitro detection of HCV  
CC antibodies in a sample, and in a serotyping assay for detecting one or  
CC more serological types of HCV present in a biological sample. The  
CC constructs can also be immobilised on a solid substrate and incorporated  
CC into a reversed phase hybridisation assay for determining the presence or  
CC the genotype of HCV. The new purification method preserves the  
CC conformation of the recombinantly expressed E1, E2 and E1/E2, and  
CC eliminates contaminating proteins. Antigens isolated using this method  
CC are more reactive with human sera than those isolated by known  
CC techniques.  
XX  
XX

Sequence 2433 BP; 434 A; 745 C; 714 G; 540 T; 0 other;

Query Match 95.7%; Score 605.8; DB 17; Length 2433;  
Best Local Similarity 98.1%; Pred. No. 3.2e-158;  
Matches 613; Conservative 0; Mismatches 12; Indels 0; Gaps 0;

QY 4 TTGGGTAAAGTATCATGATACCTTACGTGCGGCTTCGCGACCTCATGGGGTACATTCCG 63  
DB 355 TTGGGTAAAGTATCATGATACCTTACGTGCGGCTTCGCGACCTCATGGGGTACATTCCG 414  
QY 64 CTGCTGCGCGCCCTAGAGGGGTGCTGCCAGAGCCCTGGGCGATGGCGGTCTTG 123  
DB 415 CTGCTGCGCGCCCTAGAGGGGTGCTGCCAGAGCCCTGGGCGATGGCGGTCTTG 474  
QY 124 GAAGACGGGGTAAACATACACAGGAATTTGCCGTGCTTCTTCTATCTTCCTC 183  
DB 475 GAGGAGCGGGTAAACATACACAGGAATTTGCCGTGCTTCTTCTATCTTCCTC 544  
QY 184 TTGGCTTACTGCTCCTGCTGACCAATTCAGCTTCGCTTATAGGTGGCAAGCTGTC 243  
DB 535 TTGGCTTACTGCTCCTGCTGACCAATTCAGCTTCGCTTATAGGTGGCAAGCTGTC 544  
QY 244 GGGATTCACATGTCACGAACAGCTCTCCAACTCAAGCATTTGTATAGAGCGGAGC 303  
DB 595 GGGATTCACATGTCACGAACAGCTCTCCAACTCAAGCATTTGTATAGAGCGGAGC 654  
QY 304 ATGATCATGCACACCCCGGGGTGCGCTGCTGCTGCGGAGAACAACTTCCCGCTGC 363  
DB 655 ATGATCATGCACACCCCGGGGTGCGCTGCTGCTGCGGAGAACAACTTCCCGCTGC 714  
QY 364 TTGGTAGCGCTCAACCCCAAGCTCGCAGTAGGAAGCCAGCTCCCACTACGAATA 423  
DB 715 TTGGTAGCGCTCAACCCCAAGCTCGCAGTAGGAAGCCAGCTCCCACTACGAATA 774  
QY 424 CGACGCCACGTCGATTTGCTGCTGGGGGCGCTGCTTCTGTTCCGCTATGACGTGGGG 483  
DB 775 CGACGCCACGTCGATTTGCTGCTGGGGGCGCTGCTTCTGTTCCGCTATGACGTGGGG 834  
QY 484 GATCTCTGGGAGTCTGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 543  
DB 835 GATCTCTGGGAGTCTGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 894  
QY 544 GAGACGGTCAAGAGTCAATTTGCTCAATCTATCCGGCACATAACAGTACCTTATG 603  
DB 895 GAGACGGTCAAGAGTCAATTTGCTCAATCTATCCGGCACATAACAGTACCTTATG 954  
QY 604 GCTTGGGATATGATGAACGTGGT 628  
DB 955 GCTTGGGATATGATGAACGTGGT 979

RESULT 6

AAV42305  
ID AAV42305 standard; cDNA; 673 BP.

AAV42305;

XX 26-OCT-1998 (first entry)  
XX

DE HCV core protein cDNA insert of plasmid pA10.  
XX HBC; core protein; hepatitis D virus; L-HDAg; virus-like particle;  
XX infection; hepatitis B virus surface antigen; HBsAg; immunogen;  
KW vaccine; ds.  
XX  
XX Hepatitis C virus.  
OS  
XX  
XX MO9828004-A1.  
PN  
XX  
XX 02-JUL-1998.  
PD  
XX  
XX 24-DEC-1997; 97MO-AU00884.  
PF  
XX  
XX 24-DEC-1996; 96AU-0004341.  
PR  
XX  
XX (QUE-) QUEENSLAND DEPT HEALTH SAKZEMSKI VIRUS.  
PA  
XX  
XX Gowans EJ, MacNaughton TB;  
PI  
XX  
XX WPI; 1998-377411/32.  
DR  
XX  
XX P-PSDB; AAM62825.  
PT  
XX  
XX Virus-like particle for, e.g. treating microbial infection -  
PT comprises polypeptide from microorganism and sequence from Hepatitis  
PT D virus large protein, partially enveloped by Hepatitis B surface  
PT antigen  
XX  
XX Example; Fig 4; 72pp; English.  
XX  
XX This nucleotide comprises the hepatitis C virus (HCV) cDNA insert  
XX in plasmid pA10. This cDNA was obtained by PCR amplification of  
XX HCV cDNA using primers (see AAV38849-50) designed to amplify the  
XX HCV core gene. The PCR product was cloned into pBluescript KS to  
XX create pA10. It encodes a 224-amino acid polypeptide (see AAM62825).  
XX The HCV core protein can be used as an immunogen in novel fusion  
XX proteins (see AAM62657-59) that comprise HCV core protein and at  
XX least 19 amino acids (see AAM62827) of the C-terminal sequence of the  
XX large protein from hepatitis D virus (L-HDAg). In novel virus-like  
XX particles of the invention, a fusion protein immunogen is at least  
XX partially enveloped by hepatitis B surface antigen. The virus-like  
XX particle is used to ameliorate or protect against infections caused  
XX by hepatitis B virus and/or another microorganism, especially HCV.  
XX  
XX  
XX Sequence 673 BP; 115 A; 208 C; 188 G; 162 T; 0 other;

Query Match 89.2%; Score 564.8; DB 19; Length 673;  
Best Local Similarity 94.1%; Pred. No. 5.1e-147;  
Matches 587; Conservative 0; Mismatches 37; Indels 0; Gaps 0;

QY 7 GGTAAAGTATGATACCTTACGTGCGGCTTCGCGACCTCATGGGGTACATTCCGCTC 66  
DB 1 GGTAAAGTATGATACCTTACGTGCGGCTTCGCGACCTCATGGGGTACATTCCGCTC 60  
QY 67 GTGCGGCGCCCTAGAGGGGTGCTGCCAGAGCCCTGCGCATGGCGGTCTTGAA 126  
DB 61 GTGCGGCGCCCTAGAGGGGTGCTGCCAGAGCCCTGCGCATGGCGGTCTTGAG 120  
QY 127 GACGGGTAAACATACACAGGAATTTGCTGCTGCTTCTTCTATCTTCCCTTG 186  
DB 121 GACGGGTAAACATACACAGGAATTTGCTGCTGCTTCTTCTATCTTCCCTTG 180  
QY 187 GCTTACTGCTGCTGCTGACCAATTCAGCTTCGCTTATAGGTGGCAAGCTGTCGGG 246  
DB 181 GCTTACTGCTGCTGCTGCTGACCAATTCAGCTTCGCTTATAGGTGGCAAGCTGTCGGG 240  
QY 247 ATGTACCATGTCACGAACAGCTCTCCAACTCAAGCATTTGTATAGAGCGAGCAANG 306  
DB 241 GTGTACCATGTCACGAACAGCTCTCCAACTCAAGCATTTGTATAGAGCGAGCAANG 300  
QY 307 ATCATGCACACCCCGGGGTGCGCTGCTGCTGCGGAGAACAACTTCCCGCTGCTGG 366  
DB 301 ATCATGCACACCCCGGGGTGCGCTGCTGCTGCGGAGAACAACTTCCCGCTGCTGG 360

XX



QY	424	CGAGCGCACGTCGATTTGCTGTTGGGGCGGCTCTTTCTGTTCCGCTAATGTAAGTACG366G	483		
Db	1116	CGAGCGCACGTCGATTTGCTGTTGGGGCGGCTCTTTCTGTTCCGCTAATGTAAGTACG366G	1175		
OY	484	GATCTCTCGGATCTGATCTCTCTGCTCCAGCTGTTTACACATCTGCCTCGCGGCAT	543		
Db	1176	GATCTCTCGGATCTGATCTCTCTGCTCCAGCTGTTTACACATCTGCCTCGCGGCAT	1235		
OY	544	GAGACGCTGCACGACGTCGAATTGCTCAATCTATCCGCGCCACATTAACAGGTCAACCGTANG	603		
Db	1236	GAGACAGTGCAGGACGTCGAATCTCAATCTATCCGCGCCATTTATGAGTCAACCGCATG	1295		
OY	604	GCTTGGGATATGATGATGAATGTTAA	630		
Db	1296	GCTTGGGATATGATGATGAATGTTAA	1322		
RESULT 10					
ID	AAD25332	standard; cDNA; 9605 BP.			
XX	AAD25332;				
AC	12-MAR-2002	(first entry)			
DT					
XX					
DE	Hepatitis C virus (HCV) full-length cDNA mutant #2.				
XX					
KW	Hepatitis C virus; HCV; transfection; infection; virus neutralisation;				
KW	gene therapy; vaccine; immunoprotection; hepatotropic; virucide; liver;				
KW	mutant; ss.				
XX					
OS	Hepatitis C virus.				
OS	Synthetic.				
PN	MO200189364-A2.				
XX					
PD	29-NOV-2001.				
XX					
PF	23-MAY-2001; 2001WO-US16822.				
PR	23-MAY-2000; 2000US-0576989.				
XX					
PA	(UNIW ) UNIV WASHINGTON.				
XX					
PI	Rice CM, Blight KJ;				
XX					
DR	WPI: 2002-066755/09.				
XX					
PT	Hepatitis C virus variants having greater transfection efficiency and				
PT	ability to survive subpassage, useful as a vaccine for immunizing				
PT	primate to the virus, comprise non-naturally occurring viral sequences				
XX					
XX					
SS	Example 2; Page 97-100; 174pp; English.				
CC	The invention relates to Hepatitis C virus (HCV) variants which include				
CC	polynucleotides comprising non-naturally occurring HCV sequence and HCV				
CC	variants that have a transfection efficiency and ability to survive				
CC	subpassage greater than HCV that have wild-type polypeptide coding				
CC	regions. The polynucleotides of the invention are useful for identifying				
CC	a cell line that is permissive for infection with HCV and detecting				
CC	replication of HCV in cells of the cell line. They are also useful for				
CC	testing a compound for anti-viral properties and for inhibiting HCV				
CC	infection. They are also useful for the generation of defined HCV virus				
CC	stocks to develop in vitro and in vivo assays for virus neutralisation,				
CC	attachment, penetration and entry, structure/function studies on HCV				
CC	proteins and RNA elements and identification of new antiviral targets,				
CC	a systematic survey of cell culture systems and conditions to identify				
CC	those that support wild-type and variant HCV RNA replication and				
CC	particle release, production of adaptive HCV variants capable of more				
CC	efficiency replication in cell culture, production of HCV variants with				
CC	altered tissue or species tropism, establishment of alternative animal				

Query Match	Best Local Similarity	Score	DB	Length	9605;
Matches	588;	Conservative	0;	Mismatches	39;
				Indels	0;
				Gaps	0;
QY 4	TTGGGTAAAGTCATGATGATACCTTACGTACGTGCGGCTTCCGCGACCTCATGAGGGGTACATTCCG	63			
DB 696	TTGGGTAAAGTCATGATGATACCTTACGTACGTGCGGCTTCCGCGACCTCATGAGGGGTACATTCCG	755			
QY 64	CTCGTCGGGGCCCCCTTAGGGGGTCTGTGCGAAGCCCTTGGCGCANGGGGTCCGGGTTCTG	123			
DB 756	CTCGTCGGGGCCCCCTTAGGGGGTCTGTGCGAAGCCCTTGGCGCANGGGGTCCGGGTTCTG	815			
QY 124	GAGAGCGGCGTAACTATGACACAGGAGAAATTTGCTGTCTTTCTCTATCTTCTTC	183			
DB 816	GAGAGCGGCGTAACTATGACACAGGAGAAATTTGCTGTCTTTCTCTATCTTCTTC	875			
QY 184	TTGCGTTTACTGTCCTGTCTGTGACATTCACAGTTCCGCTTATGAGGTGCGCAACGTATCC	243			
DB 876	TTGCGTTTACTGTCCTGTGTGACATTCACAGTTCCGCTTATGAGGTGCGCAACGTATCC	935			
QY 244	GGGATGTACCAATGTACCAAGCAGCTGTCTCCAACTAAGCATTTGTATGAGGACAGGAC	303			
DB 936	GGGATGTACCAATGTACCAAGCAGCTGTCTCCAAAGCATTTGTATGAGGACAGGAC	995			
QY 304	ATGATCAGACACACCCCGGGGTGCGGTGCGTTCGGGAGAACAACTCTCCGGCTGC	363			
DB 996	ATGATCAGACATACCCCGGGGTGCGGTGCGTTCGGGAGAACAACTCTCCGGCTGC	1055			
QY 364	TGGGTAGGCGTACACCCCAACGCTGCGAGCTAGGAAGCGCAGGCTCCGCACTAGCAATA	423			
DB 1056	TGGGTAGGCGTACACCCCAACGCTGCGAGGCAAGAAAGGTAGGTCCCACTAGCAAGATA	1115			
QY 424	CGAGCGCACGTCGATTTGCTGTTGGGGGCGCTGCTTTCTGTGCTCCGCTATGAGTGGGG	483			
DB 1116	CGAGCGCACATGTCGATTTGCTGTTGGGGGCGCTGCTTTCTGTGCTCCGCTATGAGTGGGA	1175			
QY 484	GATTCCTGGGGGATGTCGTCCTGCTGCTGTCGCCAGCGTTTCAACATTCGCGTCGCGGGAT	543			
DB 1176	GATTCCTGGGGGATGTCGTCCTGCTGCTGTCGCCAGCGTTTCAACATTCGCGTCGCGGGAT	1235			
QY 544	GAGAGCGTGCAGAGCATGCAATTCATCTATCCCGGCAATAGACAGTCAACCGATAG	603			
DB 1236	GAGAGCATGACAGAGCATGCAATTCGTCATATATATCCCGGCAAGTACAGGTCAACCGATAG	1295			
QY 604	GCTTGGGATATGATGATGAAGTGA 630				
DB 1296	GCTTGGGATATGATGATGAAGTGA 1322				
RESULT 11					
AAD25331					
ID	AAD25331 standard; cDNA; 11062 BP.				
XX AC	AAD25331;				
XX DT	12-MAR-2002 (first entry)				
XX DE	Hepatitis C virus (HCV) full-length cDNA mutant #1.				

XX	Hepatitis C virus; HCV; transfection; infection; virus neutralisation;
KM	gene therapy; vaccine; immunoprotection; hepatotropic; virucide; liver;
XX	mutant; ss.
XX	
OS	Hepatitis C virus.
XX	Synthetic.
PM	WO200189364-A2.
PD	29-NOV-2001.
PE	23-MAY-2001; 2001WO-US16822.
PR	23-MAY-2000; 2000US-0576989.
PA	(UNIW ) UNIV WASHINGTON.
PI	Rice CM, Blight KJ;
XX	
DR	WPI; 2002-066755/09.
XX	
PT	Hepatitis C virus variants having greater transfection efficiency and
PT	ability to survive subpassage, useful as a vaccine for immunizing
PT	primate to the virus, comprise non-naturally occurring viral sequences
XX	
PS	
XX	
XX	Example 2; Page 93-96; 174pp; English.
CC	The invention relates to Hepatitis C virus (HCV) variants which include
CC	polynucleotides comprising non-naturally occurring HCV sequence and HCV
CC	variants that have a transfection efficiency and ability to survive
CC	subpassage greater than HCV that have wild-type polypeptide coding
CC	regions. The polynucleotides of the invention are useful for identifying
CC	a cell line that is permissive for infection with HCV and detecting
CC	replication of HCV in cells of the cell line. They are also useful for
CC	testing a compound for anti-viral properties and for inhibiting HCV
CC	infection. They are also useful for the generation of defined HCV virus
CC	stocks to develop in vitro and in vivo assays for virus neutralisation,
CC	attachment, penetration and entry, structure/function studies on HCV
CC	proteins and RNA elements and identification of new antiviral targets,
CC	a systematic survey of cell culture systems and conditions to identify
CC	those that support wild-type and variant HCV RNA replication and
CC	particle release, production of adaptive HCV variants capable of more
CC	efficient replication in cell culture, production of HCV variants with
CC	altered tissue or species tropism, establishment of alternative animal
CC	models for inhibitor evaluation including those supporting HCV variant
CC	replication, development of cell-free HCV replication assays,
CC	production of immunogenic HCV particles for vaccination, engineering of
CC	attenuated HCV derivatives as possible vaccine candidates, engineering
CC	of attenuated or defective HCV derivatives for expression of
CC	heterologous gene products for gene therapy and vaccine applications
CC	and for utilisation of the HCV glycoproteins for targeted delivery of
CC	therapeutic agents to the liver or other cell types with appropriate
CC	receptors. Vaccine comprising these sequences is useful for inducing
CC	immunoprotection to HCV in a primate. The present sequence is
CC	Hepatitis C virus (HCV) full-length cDNA containing the mutation
CC	that results in Ser to Ile at position 1179 of HCVrepBartzan protein,
CC	where the 5'UTR is fused to the neomycin phosphotransferase gene and
CC	the EMCV IRES is inserted upstream of the HCV open reading frame.
XX	
SO	Sequence 11062 BP; 2217 A; 3291 C; 3145 G; 2409 T; 0 other;
XX	
Query Match	89.2%; Score 564.6; DB 24; Length 11062;
Best Local Similarity	93.8%; Pred. NO. 1.5e-146;
Matches 588; Conservative	0; Mismatches 33; Indels 0; Gaps 0;
0Y	4 TTGGTAAAGTCATCATGATACCTTACGTCGGCTTCGCGACCTCATGGGTCATATTCG 63
DB	2153 TTGGTAAAGTCATCATGATACCTTACGTCGGCTTCGCGACCTCATGGGTCATATTCG 2212
0Y	64 CTCGTCGGGCGCCGCCCTAGGGGTCTGCGACAGCCCTGGGCGATGGGCTCCGGGTCTG 123

Db	2213	CTCGTCGGCGCCCCCCTAGGGGGCCCTGCGAGGGCCCTGCGCCATGAGCGCTCCGGGTTCTG	2272
Qy	124	GAAGACGGGGTGAACATATGCAACAGGGAATTTGCCGTGTGCTCTTCTATCTTCTCTC	183
Db	2273	GAGGACGGGGTGAACATATGCAACAGGGAATCTCCGGGTTGCTCTTCTATCTTCTCT	2332
Qy	184	TTGGCTTTACTCTCTGTCTGACCATTTCCAGTTCCGCTTATGAGGTGGCAAGTGTCC	243
Db	2333	TTGGCTTTGCTCTCTGTGTTGACCATATCCAGCTTCCGCTTATGAGGTGGCAAGGTATCC	2392
Qy	244	GGGATGTACCATATGTCACGAAACAGACAGCTCCCACTCAAGCATTTGTATGAGGACAGG	303
Db	2393	GGAGTGTACCATATGTCACGAAACAGACAGCTCTCCAAACGAAAGATTGTATGAGGACAGG	2452
Qy	304	ATGATCATGACACACCCCGGGTGCCTGCTCCGCTGTCGGAGAGAACACTCTCCCGCTGC	363
Db	2453	ATGATCATGACATACCCCGGGGTGCGTGCCTGCTGTCGGAGAGAACACTCTCCCGCTGC	2512
Qy	364	TGGGTAGGGCTTACCCCAACGCTCCGAGTACGAAACCCAGCGTCCCACTACGACATA	423
Db	2513	TGGGTAGGGCTTACCCCAACGCTCCGAGTACGAAACCGTACGCGTCCCACTACGAGATA	2572
Qy	424	CGAGCCACAGTGGATTTGCTGTTGGGGGGGCTGCTTCTGTTCCGCTATGTACGAGGG	483
Db	2573	CGAGCCACATGTGATTTGCTGTTGGGGGGGCTGCTTCTGTTCCGCTATGTACGAGGG	2632
Qy	484	GATCTCTGGGATCTGTCTTCCCTGCTCTCCAGCTGTTTCACCATGTCCGCTCGCGGCA	543
Db	2633	GATCTCTGGGATCTGTCTTCCCTGCTCTCCAGCTGTTTCACCATGTCTCGCTCGCGGCA	2692
Qy	544	GAGACGGTCAGAGCTGCATTTGCTCAATCTATCCCGGACACATTAACAGGTACCGTATG	603
Db	2693	GAGACAGTACAGAGCTGCATTTGCTCAATATATCCCGGACAGTACAGGTACCGTATG	2752
Qy	604	GCTTGGATATGATGATGAACGTGAA	630
Db	2753	GCTTGGATATGATGATGAACGTGCA	2779
RESULT 12			
AAA98965			
ID	AAA98965	standard; DNA; 11076 BP.	
AC	AAA98965;		
XX			
DT	08-FEB-2001	(first entry)	
XX			
DE	Hepatitis C virus DNA fragment SEQ ID NO: 1.		
XX			
KW	Cell culture; therapy; infection; vaccine; diagnosis; gene therapy; ds.		
XX			
OS	Hepatitis C virus.		
XX			
PN	DE19915178-A1.		
XX			
PD	05-OCT-2000.		
XX			
PF	03-APR-1999; 99DE-1015178.		
XX			
PR	03-APR-1999; 99DE-1015178.		
XX			
PA	(UYMA-) UNIV MAINZ GUTENBERG JOHANNES.		
XX			
PI	Bartenschlager R;		
XX			
DR	WPI; 2000-629140/61.		
XX			
PT	Cell culture system for hepatitis C virus, useful e.g. in screening for		
XX	therapeutic agents, comprises human hepatoma cells containing a viral		
PT	RNA construct that includes a selectable gene -		
XX			
PS	Claim 5; Page 14-22; 58pp; German.		
XX			

CC This invention describes a novel hepatitis C virus (HCV) cell culture  
CC system comprising human hepatoma cells that contain an integrated HCV-RNA  
CC construct (I). (I) contains the HCV-specific RNA segments 5'-NTR  
CC (non-translated region), NS (non-structural)1,3, NS4A, NS4B, NS5A, NS5B and  
CC 3'-NTR, and a selectable marker gene (II). The cell cultures, and/or  
CC (I), are used to prepare, evaluate and/or test therapeutic and/or  
CC diagnostic agents for HCV infections, and to prepare vaccines against HCV  
CC infection (particularly preparation of attenuated HCV). The can also be  
CC used for preparation of a liver-specific delivery system for gene  
CC therapy, and to identify cells permissive for HCV replication. Virus RNA  
CC replicates autonomously and with high efficiency in this cellular system,  
CC so that variations in replication rates can be measured (for screening  
CC antiviral agents) quantitatively or qualitatively using standard  
CC laboratory equipment. Efficient replication of HCV RNA is only achieved  
CC when the specified RNA segments are present and when the transfected  
CC cells are maintained under permanent selection pressure.

xx

SQ Sequence 11076 BP; 2221 A; 3297 C; 3149 G; 2409 T; 0 other;

SQ Sequence 11076 BP; 2221 A; 3297 C; 3149 G; 2409 T; 0 other;

Query Match	89.2%	Score 564.6;	DB 21;	Length 11076;
Best Local Similarity	93.8%;	Pred.No. 1.5e-146;		
Matches 588; Conservative	0;	Mismatches 39;	Indels 0;	Gaps 0;

QY	4	TTGGTAAGGTCATCATTAACCTTACGTGCGGCTTGCCGACCTCATGGGGTACATTTCCG	63
Db	2167	TTGGGTAAAGGTCATCATTAACCTTACGTGCGGCTTGCCGATTCATGGGGTACATTTCCG	2226
QY	64	CTCGTCGGCGCCCCCTAGAGGGGTGCTGGCCAGACGCCCTGGCCATAGCGCTCCGGGTTCTG	123
Db	2227	CTCGTCGGCGCCCCCTTAGGGGGGCGCTGCAGAGGCCCTTGGCCATATGCGCTCCGGGTTCTG	2286
QY	124	GAAGACGGCGGTGAACATATGCAACAGGGAATTTGCTGTGCTCTTTCTCATTTCTCCTC	183
Db	2287	GAGAGACGGCGGTGAACATATGCAACAGGGAATTTGCTGTGCTCTTTCTCATTTCTCCTT	2348
QY	184	TTGGCTTTACGTCCTATGCTGAGCAATTCACGCTTCGCTTAGAGGTGGCAAGCTGTCC	243
Db	2347	TTGGCTTTGCTGTCTTTTGGACATTCACGCTTCGCTTAGAGGTGGCAAGCTGTCC	2406
QY	244	GGGATGTACCATGTCAAGAACGACGCTCCTCAACTCAAGCATTTGTATAGGACAGCGAC	303
Db	2407	GGATGTATACCATGTCAAGAACGACGCTCCTCAAGCATTTGTATAGGACAGCGAC	2466
QY	304	ATGATCATGCACACCCCCGGGTGCGTGCCTCGGTTCGGGAACAACATCTTCCCGCTGC	363
Db	2467	ATGATCATGCATATACCCCCGGGTGCGTGCCTCGGTTCGGGAACAACATCTTCCCGCTGC	2528
QY	364	TGGGTACGCGTCACCCCCGAGCGTCGGAAGTAGGAAGCGGAGTCCGCCACTACGACATA	423
Db	2527	TGGGTACGCGTCACCCCCGAGCGTCGGAAGTAGGAAGCGGAGTCCGCCACTACGACATA	2588
QY	424	CGACGCAAGCTCGAATTTGCTGCTTTGGGGCGCGCTTCTGTTCGCGATATGACGTGGGG	483
Db	2587	CGACGCAAGCTCGAATTTGCTGCTTTGGGGCGCGCTTCTGTTCGCGATATGACGTGGGA	2648
QY	484	GATCTCTGCGGATATGTCTCTCTCTCTCTCTCCACACTTTACACATCTCGGCTGGCGGAT	543
Db	2647	GATCTCTGCGGATATGTCTCTCTCTCTCTCTCCACACTTTACACATCTCGGCTGGCGGAC	2706
QY	544	GAGACGGTGCAGACATCAATTTGCTCAATATATCCGGGACATATAACAGTCAACCGTATG	603
Db	2707	GAGACATGACAGACATCAATTTGCTCAATATATCCGGGACAGTACAGTCAACCGTATG	2766
QY	604	GCTTGGGATATGATGAAGACTGGTAA	630
Db	2767	GCTTGGGATATGATGAAGACTGGTAA	2793

RESULT 13  
AAQ24467  
ID AAQ24467 standard; DNA; 1880 BP  
XX  
AC AAQ24467;

XX	09-NOV-1992	(first entry)
DT		
XX		
DE	NANB hepatitis virus strain HC-J4 genome.	
XX		
KW	non-A, non-B hepatitis virus; NANBHV; PCR amplification	
KW	polymerase chain reaction; vaccine; antibody; ss.	
XX		
OS	Non-A, non-B hepatitis virus.	
XX		
FH	Key	Location/Qualifiers
FT	CDS	342..1880
FT		/*tag= a
FT		/label= HC-J4

CC This sentence is the genome of the non- $\beta$ , non-hepatitis virus  
CC (NANBHV) strain HC-14. This sequence was derived by amplification  
CC by polymerase chain reaction. The nucleotide sequences derived from  
CC this amplification can be used to detect NANBHV infection which could  
CC not be detected by conventional methods. The detection kits allow  
CC highly specific and sensitive detection at an early phase of  
CC infection. The polypeptide product of this coding sequence can be used  
CC for the manufacture of vaccines and immunological pharmaceuticals  
CC and also to produce antibodies specific to NANBHV.

Sequence 1880 BP; 333 A; 593 C; 556 G; 398 T; 0 other;

Query Match	88.98	Score 563	DB 13	Length 1880
Best Local Similarity	93.68	Pred. No. 2.3e-146		
Matches 587	Conservative	0	Mismatches 40	Indels 0
			Gaps	0

OY	4	TTGGGTAAAGTCAATCGATACCCCTTAACGTCGGGGCTTCGGCCAGCTCATAGGGGATACATTCGG	63
Db	636	TTGGGTAAAGTCAATCGAATACCCCTTAACGTCGGCTTCGGCTTCATCAAGGGGATATATTCGG	755
OY	64	CTCGTCGGCGCCCCCTTATGAGGGGGTGTGTCGACAGACCCCTGAGGGGATAGGCGTCCGGGCTCTG	123
Db	756	CTCGTCGGCGCCCCCTTATGAGGGGGGTGTGTCGAGAGGCCCTTTGGGACACGGGTGTCCGGGTTCTG	815
OY	124	GAAAGCGCGGTGAACATATGCAACAAGGAATTTGCCCTGGTGTCTTTTCTCTATCTTCTCTC	183
Db	816	GAGGCGCGCGTGAACATATGCAACAAGGAATTTGCCGGGTGTCTTTTCTCTATCTTCTCTC	875
OY	184	TTTGGCTTACCTGCTCTGTCTATGACATTTCCAGCTTCCGCTTATGAGGGGCCAAGCTGTCC	243
Db	876	TTTGGCTTACCTGCTCTGTCTTATGACATTTCCAGCTTCCGCTTATGAGGGGCCAAGCTGTCC	935
OY	244	GGGATGTACCATGTTCACGAACAGATGCTCTCAACTCAAGACATTTGTGTATGAGCAGCGGAC	303
Db	936	GGGATATACCATGTTCACGAACAGATGCTCTCAACTCAAGACATTTGTGTATGAGCAGCGGAC	995
OY	304	ATGATCATGCAACCCCCCGGGGTGGCTGTGCCCTTGGGGGAACAACATCTTCCCGGTGC	363

```

DB 996 ATGATCATCATACATCTCCGCGGTCGCTGCTGCTTCCGAGACAAACAGCTCCGTTGC 1055
QY 364 TGGAGCGCTCACCCCAAGCTCGAGCTAGAGAACGCCAGGCTCCCACTACGACAAATA 423
DB 1056 TGGAGTGGCTCTACTCCCAAGCTCGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1115
QY 424 CGACGCGACGCTGATTTGCTGCTGCGGCGGCTCTTTGTTGCTGCTGCTGCTGCTGCTG 483
DB 1116 CGACGCGACGCTGATTTGCTGCTGCGGCGGCTCTTTGTTGCTGCTGCTGCTGCTGCTG 1175
QY 484 GATCTCTGCGATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 543
DB 1176 GATCTCTGCGATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1235
QY 544 GAGACGCTGACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 603
DB 1236 GAGACGCTGACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1295
QY 604 GCTTGGGATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 630
DB 1296 GCTTGGGATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1322

```

## RESULT 14

AAQ20923 standard; DNA; 932 BP.

```

AC AAQ20923;
XX
DT 06-MAY-1992 (first entry)
XX
DE C10-E12 DNA fragment encoding NANBH-specific antigen polypeptide.
XX
KM Non-A non-B hepatitis virus; recombinant; detection; ss.
XX
OS Non-A non-B hepatitis virus.
XX
FH Key Location/Qualifiers
FT CDS 3..932
FT tag a
XX
PN EP468657-A.
XX
PD 29-JAN-1992.
XX
PE 08-JUL-1991; 91EP-0306158.
XX
PR 20-DEC-1990; 90JP-0413844.
PR 09-JUL-1990; 90JP-0180889.
PR 30-NOV-1990; 90JP-0339589.
XX
PA (TOFU ) TONEN CORP.
XX
PI Maki N, Yamaguchi K, Toyoshima A, Kohara M;
XX
DR WPI; 1992-034390/05.
DR P-PSDB; AAR20720.
XX
PT Non-A, non-B hepatitis-specific antigen polypeptide - for detection
PT of hepatitis virus gene or antibody directed against virus
XX
PS Claim 15; Fig 15; 78bp; English.
XX
CC The sequence is that of a DNA fragment encoding a non-A non-B (NANB)
CC hepatitis-specific antigen polypeptide, it can be used to produce the
CC polypeptide recombinantly. It is derived from the DNA clone C10-E12.
CC It can be used to give PCR primers which are capable of detecting NANB
CC hepatitis with high accuracy. See also AAQ20617-Q20629 and
CC AAQ20922-Q20926.
XX
SQ Sequence 932 BP; 173 A; 276 C; 257 G; 226 T; 0 other;

```

Query Match 88.7%; Score 561.4; DB 13; Length 932;  
 Best Local Similarity 93.5%; Pred. No. 5e-146;  
 Matches 586; Conservative 0; Mismatches 41; Indels 0; Gaps 0;

```

QY 4 TTGGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 63
DB 9 TTGGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 68
QY 64 CTGCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 123
DB 69 CTGCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 128
QY 124 GAGACGCGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 183
DB 129 GAGACGCGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 188
QY 184 TTGGGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 243
DB 189 TTGGGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 248
QY 244 GGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 303
DB 249 GGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 308
QY 304 ATGATCATGACACACCCCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 363
DB 309 ATGATCATGACACACCCCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 368
QY 364 TGGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 423
DB 369 TGGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 428
QY 424 CGACGCGACGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 483
DB 429 CGACGCGACGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 488
QY 484 GATCTCTGCGATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 543
DB 489 GATCTCTGCGATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 548
QY 544 GAGACGCTGACGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 603
DB 549 GAGACGCTGACGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 608
QY 604 GCTTGGGATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 630
DB 609 GCTTGGGATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 635

```

## RESULT 15

AAQ29628 standard; DNA; 2540 BP.

```

AC AAQ29628;
XX
DT 16-MAR-1993 (first entry)
XX
DE Hepatitis C virus HC-J4 5' region.
XX
KM Non-A non-B hepatitis; NANBH; HCV; detection; diagnosis; screening;
KM PCR; primer: polymerase chain reaction; ss.
XX
OS Hepatitis C virus.
XX
PN EP510952-A.
XX
PD 28-OCT-1992.
XX
PF 23-APR-1992; 92EP-0303625.
XX
PR 26-APR-1991; 91JP-0191376.
XX
PA (IMMO ) IMMUNO JAPAN INC.

```



GenCore version 5.1.3  
Copyright (c) 1993 - 2003 Compugen Ltd.

## OM nucleic - nucleic search, using sw model

Run on: February 18, 2003, 05:52:15 ; Search time 1691 Seconds  
(without alignments)  
6062.533 Million cell updates/sec

Title: us-09-899-303a-7

Perfect score: 633  
Sequence: 1 ATGTTGGGTAGGTCATCGA.....TGATGATGAACGTGTAATAG 633

## Scoring table:

IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%  
Listing first 45 summaries

## Database :

EST:\*  
1: em\_estb:\*  
2: em\_esthum:\*  
3: em\_estin:\*  
4: em\_estnu:\*  
5: em\_estov:\*  
6: em\_estpl:\*  
7: em\_estro:\*  
8: em\_hlc:\*  
9: gb\_est1:\*  
10: gb\_est2:\*  
11: gb\_hlc:\*  
12: gb\_est3:\*  
13: gb\_est4:\*  
14: gb\_est5:\*  
15: em\_estfun:\*  
16: em\_estom:\*  
17: gb\_gss:\*  
18: em\_gss\_hum:\*  
19: em\_gss\_inv:\*  
20: em\_gss\_pln:\*  
21: em\_gss\_vtc:\*  
22: em\_gss\_fun:\*  
23: em\_gss\_man:\*  
24: em\_gss\_mus:\*  
25: em\_gss\_other:\*  
26: em\_gss\_pro:\*  
27: em\_gss\_rtd:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	75.4	11.9	488	10	AV755731
C 2	61	9.6	492	10	AV758366
C 3	39.8	6.3	1270	13	BG968359
C 4	39.6	6.3	423	9	AA459034
C 5	39	6.2	359	13	BJ252669
C 6	39	6.2	375	13	BJ246716

7	38.6	6.1	399	10	AV638521	AV638521
8	38.6	6.1	434	10	AV637507	AV637507
9	38.6	6.1	440	10	AV637983	AV637983
10	38.6	6.1	450	10	AV637259	AV637259
11	38.6	6.1	451	10	AV637328	AV637328
12	38.6	6.1	451	10	AV637643	AV637643
13	38.6	6.1	453	10	AV634724	AV634724
14	38.6	6.1	454	10	AV637050	AV637050
15	38.6	6.1	456	10	AV635382	AV635382
16	38.6	6.1	473	10	AV632765	AV632765
17	38.6	6.1	481	10	AV635503	AV635503
18	38.6	6.1	485	10	AV632811	AV632811
19	38.6	6.1	506	10	AV632445	AV632445
20	38.6	6.1	508	10	AV634095	AV634095
21	38.6	6.1	526	10	AV641895	AV641895
22	38.6	6.1	533	10	AV638125	AV638125
23	38.6	6.1	537	10	AV632335	AV632335
24	38.6	6.1	588	10	AV387329	AV387329
25	38.6	6.1	525	10	BE337089	BE337089
26	38.4	6.1	671	13	BI723733	BI723733
27	38.2	6.0	431	10	AV639153	AV639153
28	38.2	6.0	501	10	AV638474	AV638474
29	38	6.0	1195	17	BI09802	BI09802
30	37.8	6.0	435	14	C72860	C72860
31	37.6	5.9	222	13	BM256206	BM256206
32	37.6	5.9	456	13	BI776179	BI776179
33	37.6	5.9	478	12	BF775654	BF775654
34	37.6	5.9	486	13	BI776135	BI776135
35	37.6	5.9	570	13	BI776136	BI776136
36	37.6	5.9	593	13	BI849151	BI849151
37	37.4	5.9	497	10	AV633658	AV633658
38	37.4	5.9	624	13	BI723734	BI723734
39	37.2	5.9	925	17	CNS0091P	CNS0091P
40	37	5.8	945	17	CNS05KMC	CNS05KMC
41	37	5.8	970	17	CNS010C9	CNS010C9
42	36.8	5.8	431	10	AV636681	AV636681
43	36.8	5.8	439	10	AV638045	AV638045
44	36.8	5.8	477	10	AV635762	AV635762
45	36.8	5.8	504	10	AV637539	AV637539

## ALIGNMENTS

RESULT 1  
LOCUS AV755731/c 488 bp mRNA linear EST 19-OCT-2000  
DEFINITION AV755731 BM Homo sapiens CDNA clone BMRK03 5', mRNA sequence.  
ACCESSION AV755731  
VERSION AV755731.1 GI:10913579  
KEYWORDS EST.

## SOURCE

human.

## ORGANISM

Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

## REFERENCE

1 (bases 1 to 488)  
Gu,J., Zhao,M., Huang,Q., Xu,X., Li,Y., Peng,Y., Song,H., Xiao,H.,  
L., Xu,S., Gu,W., Tu,Y., Jia,J., Fu,G., Ren,S., Zhong,H., Lu,G.,  
Yang,Y., Gao,G., Wang,Z., Zhang,Q., Chen,S., Han,Z. and Chen,Z.

## TITLE

Homo sapiens CDNA BM clones

## JOURNAL

Unpublished (2000)

## COMMENT

Contact: Zeguang Han  
Chinese National Human Genome Center at Shanghai  
351 Guo Shoujing Road, Zhangjiang Hi-Tech Park, Pudong, Shanghai  
201203, P. R. China  
Tel: 86-21-50801912(ex.45)  
Fax: 86-21-50801922  
Email: hanzg@chgc.sh.cn  
This clone is available at CHGC in Shanghai.

## FEATURES

source

location/Qualifiers  
1..488  
/organism="Homo sapiens"

```

/db_xref="taxon:9606"
/clone="BMFAK03"
/clone_lib="BM"
/tissue_type="Bone marrow"
/lab_host="CD34+ hematopoietic stem/progenitor cell"
/note="vector: pTriplex2; Site_1: sfl1a; Site_2: sfl1B"
BASE COUNT      116 a      134 c      137 g      97 t      4 others
ORIGIN

Query Match      11.9%; Score 75.4; DB 10; Length 488;
Best Local Similarity 70.3%; Pred. No. 1.7e-09;
Matches 130; Conservative 0; Mismatches 51; Indels 4; Gaps 2;

QY 445 GTTGGGGCGGCTCTTCTGTCGCTATGATGATGAGATCTGCGGATCTGCTTC 504
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 472 GTGATGACACATGCTCTGCTCAGCTCTGATGATGAGATCTGCGGATGATG 413
QY 505 CTGCTGCCAGCTGTTCACCATCTGCGCTCGCGGATGAGACGTCGAGCTGCAAT 564
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 412 CTGCGACATTCAGCTG--ATCATCTGCGCTCAGCACCATGATGTTGTCATGATGCAAC 356
QY 565 TGCATATCTATCCGCGCCACATACAGTCAACGATG--GCTTGGGATATGATGATGAA 623
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 355 TGCTCATCTATCTCTGCGGCATCATGACACCGTATGAGCATGGACATGATGATGAA 296
QY 624 CTGGT 628
    ||| |||
DB 295 CTGGT 291

RESULT 2
AV758366/c      492 bp      mRNA      linear      EST 19-OCT-2000
LOCUS      AV758366 BM Homo sapiens CDNA clone BMFAK03 5', mRNA sequence.
DEFINITION      AV758366
ACCESSION      AV758366
VERSION      AV758366.1 GI:10916214
KEYWORDS      EST.
SOURCE      human.
ORGANISM      Homo sapiens
                Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
                Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE      1 (bases 1 to 492)
AUTHORS      Gu,J., Zhao,M., Huang,Q., Xu,X., Li,Y., Peng,Y., Song,H., Xiao,H.,
                Gu,Y., Li,N., Qian,B., Liu,F., Qu,J., Gao,X., Cheng,Z., Xu,Z., Zeng
                L., Xu,S., Gu,W., Tu,Y., Jia,D., Fu,G., Ren,S., Zhong,M., Lu,G.,
                Yang,Y., Gao,G., Wang,Z., Zhang,Q., Chen,S., Han,Z. and Chen,Z.
                Homo sapiens CDNA BM clones
                Unpublished (2000)
TITLE      Contact: Zeguang Han
JOURNAL      Chinese National Human Genome Center at Shanghai
                351 Guo Shoujing Road, Zhangjiang Hi-Tech Park, Pudong, Shanghai
                201203, P. R. China
                Tel: 86-21-50801919(ex.45)
                Fax: 86-21-50801922
                Email: hanzg@hgc.sh.cn
                This clone is available at CHGC in Shanghai.
FEATURES
    source
        1..492
        /organism="Homo sapiens"
        /db_xref="taxon:9606"
        /clone="BMFAK03"
        /clone_lib="BM"
        /tissue_type="Bone marrow"
        /cell_type="CD34+ hematopoietic stem/progenitor cell"
        /lab_host="BM25.8"
        /note="vector: pTriplex2; Site_1: sfl1a; Site_2: sfl1B"
BASE COUNT      124 a      128 c      125 g      112 t
ORIGIN

Query Match      9.6%; Score 61; DB 10; Length 492;
Best Local Similarity 68.0%; Pred. No. 1.1e-05;
Matches 115; Conservative 0; Mismatches 50; Indels 4; Gaps 2;

```

```

QY 461 TCTGTTCCGCTATGATGAGTGGGATCTCTGCGGATCTGCTTCCTCCACCTGT 520
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 457 TGTGATCAGCTACTACAGTGTGACCTCTGCTGGGATGCTTGACGCCAAGCTG- 399
QY 521 TCACCATCTGCTCGCGGCATGAGACGGTGTGAGACATGCAATTCCTAATCTACCCG 580
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 398 -ATTATCTCTGACAGCAACATTTGTTGTGCAAAAGCACTGCTCATCTTACCTG 341
QY 581 GCCCATTAACAGGTC-ACCGTATGCTTGGGATGATGATGATGATGATGATGAT 628
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 340 GCTGCACTACATGACATACATGATGATGATGATGATGATGATGATGATGATGAT 292

RESULT 3
BG968359/c      1270 bp      mRNA      linear      EST 12-JUN-2001
LOCUS      602834809F1 NCL_CGAP_Co24 Mus musculus CDNA clone IMAGE:4989195 5',
DEFINITION      mRNA sequence.
ACCESSION      BG968359
VERSION      BG968359.1 GI:14355996
KEYWORDS      EST.
SOURCE      house mouse.
ORGANISM      Mus musculus
                Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
                Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE      1 (bases 1 to 1270)
AUTHORS      NIH-MGC http://mgc.nci.nih.gov/.
TITLE      Mammalia: Eutheria: Rodentia; Sciurognathi; Muridae; Murinae; Mus.
JOURNAL      National Institutes of Health, Mammalian Gene Collection (MGC)
                Unpublished (1999)
COMMENT      Contact: Robert Strausberg, Ph.D.
                Email: cgabbs-remail.nih.gov
                Tissue Procurement: Jeffrey E. Green, M.D.
                CDNA Library Preparation: Life Technologies, Inc.
                DNA library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
                DNA sequencing by: Incyte Genomics, Inc.
                Clone distribution: MGC clone distribution information can be
                found through the I.M.A.G.E. Consortium/LLNL at:
                http://image.llnl.gov
                Plate: L1AM1003 row: a column: 04
                High quality sequence start: 6
                High quality sequence stop: 416.
FEATURES
    source
        1..1270
        /organism="Mus musculus"
        /strain="FVB/N"
        /db_xref="taxon:10090"
        /clone="IMAGE:4989195"
        /clone_lib="NCL_CGAP_Co24"
        /lab_host="DH10B (T1 phage-resistant)"
        /note="Organ: colon; Vector: PCMV-SPORT6; Site_1: NotI;
                Site_2: SalI; Cloned unidirectionally. Primer: Oligo dT.
                Average insert size 1.6 kb. Constructed by Life
                Technologies. Note: this is a NCL_CGAP Library."
BASE COUNT      274 a      387 c      445 g      164 t
ORIGIN

Query Match      6.3%; Score 39.8; DB 13; Length 1270;
Best Local Similarity 45.0%; Pred. No. 6.2;
Matches 149; Conservative 0; Mismatches 182; Indels 0; Gaps 0;

```

OY 261 GAACACGCGTCACCACTCAAGCATTTGTGATAGGCAAGGAGCATGATCATGACACACCCC 320  
 Db 945 CATCCCCCTCGTGTATCTTATATAGGCGCTCCGTCGCCCTGCCCGCCGCTTTTCCCATGCGC 886  
 OY 321 CGGGTGCCTGCCTCGTTCCGGAGAACAACTCTTCCGCGTCTGGGTAGCGCTACCCC 380  
 Db 885 CGGGCTCGGCTCAAGGCTCTCTGTGACGTTCGTCTGCTGCGCCCGGGGCGCTGTCAAGC 826  
 OY 381 CACGCTGCAGCTAGGAACGCCAGGCTCCCC 411  
 Db 825 TTGCTTCGCGCTCTTGGGGTACCTCTTCCGC 795

RESULT 4					
AA459034					
LOCUS	AA459034	423 bp	mRNA	linear	EST 13-AUG-1997
DEFINITION	aa26b11.1 NCI_CGAP_GCB1 Homo sapiens cDNA clone IMAGE:814365 5',				
	mRNA sequence.				

VERSION	AA459034.1	GI:2183941
KEYWORDS	EST,	
SOURCE	human.	
ORGANISM	Homo sapiens	

REFERENCE  
AUTHORS  
TITLE

1 (bases 1 to 423)  
NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.  
National Cancer Institute, Cancer Genome Anatomy Project (CGAP).  
Mammalia; Eutheria; Primates; Cercarihinii; Homnidae; Homo.  
Euarchyotia; Metazoa; Chordata; Crniata; Vertebrati; Euteleostomi;  
National Cancer Institute, Cancer Genome Anatomy Project (CGAP).

**JOURNAL COMMENT** Unpublished (1997)  
Contact: Robert Strausberg, Ph.D.

This clone is available royalty-free through LLNL; contact the IMAGE Consortium ([info@image.llnl.gov](mailto:info@image.llnl.gov)) for further information. Seq primer: -28ml3 rev2 ET from Amersham High quality sequence stop: 382.

FEATURES	Location/Qualifiers
source	1. . 423

```

/organism="Homo sapiens"
/db_xref="GDB:6032522"
/db_xref="taxon:9606"
/clone_1 IMAGE:814365"
/clone_1ID="NCI_CGAP_GCB1"
/tissue_type="germinal center B cell"
/lab_host="DH10B"
/Note="Vector: pRTT3D-Pac (Pharmacia) with a modified
polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
was prepared from human tonsillar cells enriched for
germinal center B cells by flow sorting (CD20+, IgD-),
provided by Dr. Louis M. Staudt (NCI), Dr. David Allman
(NCI) and Dr. Gerald Marti (CBER). cDNA synthesis was
primed with a Not I - oligo(dT) primer
[5'-TGTTACCAATCTGAGATGGAGGCGCCCTCATTTTTTTTTTTTTTTT-3'
]. Double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Not I and cloned into the Not I
and Eco RI sites of the modified pRTT3 vector. Library
went through one round of normalization, and was
constructed by Bento Soares and M. Fatima Bonaldo."

```

	Query Match	Best Local Similarity	Score 39.6	DB 9	Length 423
	Matches 114	Conservative	47.9%	Pred. No. 5.1	
			0	Mismatches 124	Indels 0; Gaps 0;
QY	348	CAACCTTTCCCGCTCTGTGGTTCAGCCCTCAACCCCAACGCTCGACGCTAGGAGAACCCAGCGT	407		
DB	83	CATGCTGCTGCTGCTGCGCGGAGCCACCGCTCTCGGAGACAGGAGACACTCCCTTACACA	142		
QY	408	CCCCACTAGCAATAATAGACAGCCACGCTGATTTGCTGTTGGGGCGGCTGCTTTCTGTTTC	467		
DB	143	CCCTAGCCCTTCGCGCTTCGCGGCGCCCTATCTCTCTTCCTGCTCTGTGCTCCCTACTATTC	202		

0y GCCTATTACGTCGGGGGATCTCTGGGATCTGTCTTCTCGTCACAGCTGTTACCAT 527  
468  
Db CATCAGGGAGCAGCTGACTTACAGGATCCCCGACGACCTGGCTAGACATTACAG 262  
203  
0y CTCGCTCGCGCGATGAGACGGTGCAGACATGCATATGCTCATCTTATCCCCGGCAC 585  
528  
Db CACGTCCTTCCAGCTGAGCCAGCGCCAGGTTGGGCAAGGGGCTTCTGCGCCCCCCCC 320  
263

RESULT 5	LOCUS	DEFINITION	BU252669/c	BU252669	359 bp	mRNA	linear	EST 05-APR-2002
				BU252669 Y. Ogihara unpublished cDNA library, <i>Wb f</i> <i>Triticum aestivum</i> cDNA clone wht25g19 3', mRNA sequence.				

VERSION	BJ252669.1	GI:20061830
KEYWORDS	EST.	
SOURCE	bread wheat.	
ORGANISM	Triticum aestivum	

Euaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooidae;  
; Triticeae; Triticum.  
1 (bases 1 to 359)

**AUTHORS** Ogihara, Y. and Mural, R.  
**TITLE** Expressed genes in *Triticum aestivum*  
**JOURNAL** Unpublished (2002)  
**COMMENT** Contact: Tadasu Shin-1  
Center For Genetic Resource Information

1111 Yata, Mishima, Shizuoka 411-8540, Japan  
Tel.: 81-559-81-6856  
Fax: 81-559-81-6855  
Email: tshin@genes.nig.ac.jp.

FEATURES	Location/Qualifiers
source	1. .359

/organism="Triticum aestivum"  
 /cultivar="Chinese Spring"  
 /db\_xref="taxon:4565"  
 /clone="whf25g19"  
 /clone\_id="Y. Ogihara unpublished cDNA library, wh\_f"  
 /tissue\_type="spike at flowering state"  
 /dev\_stage="Peekes', scale 10.5.1"  
 /note="Vector: lambda Uni-ZAP XR, excised phagemid;  
 Site\_1: EcoRI; Site\_2: XhoI; Plants were grown under  
 hydroponic conditions at UC Davis, salt stressed for 12  
 hours, and for 7 days, then dissected and frozen (Akhunov  
 in J Dvorak lab). Total RNA was prepared from sheath  
 tissue, equal quantities of RNA were pooled from the two  
 samples, polyA was purified from the pooled RNA, a cDNA  
 library was made, and the cDNA clones were in vivo  
 excised to give plasmid phagemids in the Tj Close lab  
 at the University of California, Riverside (Akhunov, Chin  
 , Choi, Close, Fenton, Kianian, Otto, Simons, Zhang).  
 Plasmid DNA preparations and DNA sequencing were  
 performed in the OD Anderson lab (all other authors)."

```
.RESULT 6
BU245716

Query Match          6.2%; Score 39; DB 13; Length 359;
Best Local Similarity 58.0%; Pred. No. 7;
Matches 69; Conservative 0; Mismatches 50; Indels 0; Gaps 0;
```

LOCUS	BJ246716	375 bp	mRNA	linear	EST 05-APR-2002
DEFINITION	BJ246716 Y. Ogihara unpublished cDNA library, wh_f Triticum aestivum cDNA clone wh25919 5', mRNA sequence.				
ACCESSION	BJ246716				
VERSION	BJ246716.1	GI:20058228	EST.		
KEYWORDS	bread wheat.				
SOURCE	Triticum aestivum				
ORGANISM	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooidaeae; Triticeae; Triticum.				
REFERENCE	1 (bases 1 to 375)				
AUTHORS	Ogihara,Y. and Murai,K.				
TITLE	Expressed genes in Triticum aestivum				
JOURNAL	Unpublished (2002)				
COMMENT	Contact: Tadasu Shin-I Center For Genetic Resource Information National Institute of Genetics 1111 Yata, Mishima, Shizuoka 411-8540, Japan Tel: 81-559-81-6856 Fax: 81-559-81-6835 Email: tshini@genes.nig.ac.jp. Location/Qualifiers				
FEATURES	source				
	1..375				
	/organism="Triticum aestivum"				
	/cultivar="Chinese Spring"				
	/db_xref="taxon:4563"				
	/clone="wh25919"				
	/clone_lib="Y. Ogihara unpublished cDNA library, wh_f"				
	/tissue_type="spike at flowering date"				
	/dev_stage="Peekes' scale 10.5.1"				
	/note="Vector: lambda Uni-ZAP XR, excised phagmid; Site_1: EcoRI; Site_2: XhoI; Plants were grown under hydroponic conditions at UC Davis, salt stressed for 12 hours, and for 7 days, then dissected and frozen (Akhunov in J Dvorak lab). Total RNA was prepared from sheath tissue, equal quantities of RNA were pooled from the two samples, polyA was purified from the pooled RNA, a cDNA library was made, and the cDNA clones were in vivo excised to give pluescript phagemids in the TV Close lab at the University of California, Riverside (Akhunov, Chin, Choi, Close, Fenton, Kianian, Otto, Simons, Zhang). Plasmid DNA preparations and DNA sequencing were performed in the OD Anderson lab (all other authors)."				
BASE COUNT	81 a	107 c	110 g	77 t	
ORIGIN					
Query Match	6.2%; Score 39; DB 13; Length 375;				
Best Local Similarity	58.0%; Pred. No. 7.1;				
Matches	69; Conservative	0; Mismatches	50; Indels	0; Gaps	0;
Oy	24 CCTTACGTGCGGCTTCGGCAGCCATCATGGGTTACATTCCGTCGTGCAGCCCCCTTAGG	83			
Dd	36 CTTCGAATGCAACACAGCCGCCGCTGTGGAAGGCCCTCAAGGGGGGTGCAAGCCGCTGCGTGCG	95			
Oy	84 GGCTCTGCGCACAGCCCCTGGCGCCATGCGGCTCCGGGTTCTGCAAGACGGCGTAACATANG	142			
Dd	96 GGACGCGGACAGCCCTGGGCGCCAGAGAGCTGACACTGCTGCGCGGTGACAGTCCCAAGG	154			
RESULT 7					
LOCUS	AV638521	399 bp	mRNA	linear	EST 15-DEC-2000
DEFINITION	AV638521 Chlamydomonas reinhardtii 5% CO2 chlamydomonas reinhardtii				
ACCESSION	AV638521				
KEYWORDS	AV638521.1	GI:10781841	EST.		
SOURCE	Chlamydomonas reinhardtii				
ORGANISM	Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales; Chlamydomonadaeae; Chlamydomonas.				
REFERENCE	1 (bases 1 to 399)				

```

AUTHORS      Asamizu,E., Miura,K., Kuchio,K., Inoue,Y., Fukuzawa,H., Ohyama,K.,
              Nakamura,Y. and Tabata,S.
TITLE        Generation of expressed sequence tags from low-CO2 and high-CO2
              adapted cells of Chlamydomonas reinhardtii
JOURNAL      DNA Res. 7 (5), 305-307 (2000)
MEDLINE      20539644
COMMENT      Contact: Erika Asamizu
              The First Laboratory for Plant Gene Research
              Kazusa DNA Research Institute
              Yana 1532-3, Kisarazu, Chiba 292-0812, Japan
              Email: asamizue@kazusa.or.jp, URL: http://www.kazusa.or.jp/en/plant/.
FEATURES
SOURCE       1. .399
              /organism="Chlamydomonas reinhardtii"
              /strain="C9"
              /db_xref="taxon:3055"
              /clone="HC087407_-x"
              /note="lib="Chlamydomonas reinhardtii 5% CO2"
              /note="Vector: plusescriptII SK-; Site_1: EcoRI; Site_2:
              XhoI; The cDNA library was constructed from cells cultured
              in a medium with bubbling air containing 5% carbon
              dioxide"
BASE COUNT   70 a      144 c      122 g      63 t
ORIGIN
Query Match   6.1%; Score 38.6; DB 10; Length 399;
Best Local Similarity 50.8%; Pred. No. 9.3;
Matches 92; Conservative 0; Mismatches 89; Indels 0; Gaps 0;
QY 183 CTTGGCTTTACTGTCGCTCTGCAGCAATTCGACGCTCCGCTATGAGGCGCAACGCTGC 242
Db 44 CTTGCACCCCTTGACGCGCTGCTGCATGCTGCACACCAACTTCGCGGTGGACCATCTT 103
QY 243 CGGGATGTACCATGTCACAGCAAGACTGCTCCAACTCAAGCATTTGTATGAGCAGCGGA 302
Db 104 CGGCGTGTGGCCCGGCGAGCAAGCTGACCAATACACGCGCGCGAGCAGGTGGTGGCGG 163
QY 303 CATGATCATGCACACACCCCGGGTGGTCCCTTCGCTTGGGGAAGAACTTCTCCCGTG 362
Db 164 CATGGGCATCTACGCGTCCCGCAGCAGCTGTTCGATTCGCTGAGGAGAGCAGCGCCGCGTG 223
QY 363 C 363
Db 224 C 224

RESULT 8
AV637507 434 bp mRNA linear EST 15-DEC-2000
LOCUS      AV637507 Chlamydomonas reinhardtii 5% CO2 Chlamydomonas reinhardtii
DEFINITION cDNA clone HC074a01_r 5', mRNA sequence.
ACCESSION  AV637507
VERSION     AV637507
KEYWORDS
SOURCE      AV637507.1 GI:10780827
ORGANISM    Esf.
            Chlamydomonas reinhardtii.
            Chlamydomonas reinhardtii.
            Chlamydomonas reinhardtii.
            Chlamydomonas reinhardtii.
            Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;
            Chlamydomonadaceae; Chlamydomonas.
REFERENCE   1 (bases 1 to 434)
            Asamizu,E., Miura,K., Kuchio,K., Inoue,Y., Fukuzawa,H., Ohyama,K.,
            Nakamura,Y. and Tabata,S.
            Generation of expressed sequence tags from low-CO2 and high-CO2
            adapted cells of Chlamydomonas reinhardtii
            DNA Res. 7 (5), 305-307 (2000)
COMMENT     Contact: Erika Asamizu
            The First Laboratory for Plant Gene Research
            Kazusa DNA Research Institute
            Yana 1532-3, Kisarazu, Chiba 292-0812, Japan
            Email: asamizue@kazusa.or.jp, URL: http://www.kazusa.or.jp/en/plant/.
FEATURES
SOURCE      1. .434
            /organism="Chlamydomonas reinhardtii"

```

BASE COUNT		80 a	149 c	130 g	75 t
ORIGIN					
Query Match					
Best Local Similarity 50.8%; Pred. No. 9.5; Mismatches 89; Indels 0; Gaps 0;					
Matches 92; Conservative 0;					
OY	183	CTTGGCTTTACTGTCCTGTCTGACATTCACGCTTCGCTTATGAGGTGGCGAACGTGTC	242	/straln="CG" /db.xref="taxon:3055" /clone_id="HC074801_r" /note="Vector: plasmid p153-; Site 1: EcoRI; Site 2: XhoI; The CDNA library was constructed from cells cultured in a medium with bubbling air containing 5% carbon dioxide"	
Db	21	CTTGCACCCCGGACGGCTGTCATGTCGACACCAACTTCGCGGTGGGACCATCTT	80		
OY	243	CGGAGTGTACCATGTCTACGAGACGACTGCTCAACTCAAGCATTTGTATGAGCGACGGCA	302		
Db	81	CGGCGTGTGGCGCGCGGACCAACGTCACCACTCCGCGCGGAGGTGGCGCGG	140		
OY	303	CATGATCATGTGACACCCCGGGGTGGCGCTTCGCTGGGAGAACATCTTCCCGTG	362		
Db	141	CATGGGCACTTCAAGGTCCCGGACCGCTGTCTGTGCAATTCCTGAAAGACGCCCGGCTG	200		
OY	363	C	363		
Db	201	C	201		
RESULT 9					
LOCUS AV637983					
DEFINITION AV637983 Chlamydomonas reinhardtii 5% CO2 Chlamydomonas reinhardtii					
ACCESSION AV637983					
VERSION AV637983.1					
KEYWORDS GI:10781303					
SOURCE EST.					
ORGANISM Chlamydomonas reinhardtii.					
Chlamydomonas reinhardtii					
Eukaryote; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;					
Chlamydomonadaceae; Chlamydomonas.					
REFERENCE 1 (bases 1 to 440)					
AUTHORS Asamizu,E., Miura,K., Kuchio,K., Inoue,Y., Fukuzawa,H., Ohyama,K., Nakamura,Y. and Tabata,S.					
TITLE Generation of expressed sequence tags from low-CO2 and high-CO2 adapted cells of Chlamydomonas reinhardtii					
JOURNAL DNA Res. 7 (5), 305-307 (2000)					
MEDLINE 20539644					
COMMENT Contact: Erika Asamizu					
The First Laboratory for Plant Gene Research					
Kazusa DNA Research Institute					
Yana 1532-3, Kisarazu, Chiba 292-0812, Japan					
Email: asamizu@kazusa.or.jp, URL: http://www.kazusa.or.jp/en/plant/.					
FEATURES					
source					
1..440					
/organism="Chlamydomonas reinhardtii"					
/strain="CG"					
/db_xref="taxon:3055"					
/clone_id="HC080C04_r"					
/note="Vector: plasmid p153-; Site 1: EcoRI; Site 2: XhoI; The CDNA library was constructed from cells cultured in a medium with bubbling air containing 5% carbon dioxide"					
BASE COUNT					
ORIGIN					
Query Match					
Best Local Similarity 50.8%; Pred. No. 9.5; Mismatches 89; Indels 0; Gaps 0;					
Matches 92; Conservative 0;					
OY	183	CTTGGCTTTACTGTCCTGTCTGACATTCACGCTTCGCTTATGAGGTGGCGAACGTGTC	242	/straln="CG" /db.xref="taxon:3055" /clone_id="HC074801_r" /note="Vector: plasmid p153-; Site 1: EcoRI; Site 2: XhoI; The CDNA library was constructed from cells cultured in a medium with bubbling air containing 5% carbon dioxide"	
Db	21	CTTGCACCCCGGACGGCTGTCATGTCGACACCAACTTCGCGGTGGGACCATCTT	80		
OY	243	CGGAGTGTACCATGTCTACGAGACGACTGCTCAACTCAAGCATTTGTATGAGCGACGGCA	302		
Db	81	CGGCGTGTGGCGCGCGGACCAACGTCACCACTCCGCGCGGAGGTGGCGCGG	140		
OY	303	CATGATCATGTGACACCCCGGGGTGGCGCTTCGCTGGGAGAACATCTTCCCGTG	362		
Db	141	CATGGGCACTTCAAGGTCCCGGACCGCTGTCTGTGCAATTCCTGAAAGACGCCCGGCTG	200		
OY	363	C	363		
Db	201	C	201		

OY	183	CTTGCCATTACTGCTGTGCATGACATTCCAGCCTTCGGTTATGAGGGCGCAAGTGTG	242
Dd	80	CTTGCACCCCCCTGGACGGCTGCTCCATGTGTGACACCACATTCGCCGTGGCACCATCTT	139
OY	243	CGGGATGTACCATGTGCAGGAAGCACTGCTCCAACTCAAGCATTTGTATGAGGACGCGA	302
Dd	140	CGGGCTGTGGCCCGCGCAACAACCTGACCAAATFACACGGCCGCGAGAGGTGGTGGCGG	199
OY	303	CATGATCATGCAACACCCCCGGTGGTGGCTTCCGCTTGGGAGAACAACCTTTCGGCTG	362
Dd	200	CATGGGCACTTAAGGTCCCGCAGCAGTGTTCGATTGCTCCCTGAAGAGAGACCCCGGCTG	259
OY	363	c	363
Dd	260	c	260
RESULT	10		
LOCUS	AV637259	450 bp	mRNA linear EST 15-DEC-2000
DEFINITION	AV637259 Chlamydomonas reinhardtii 58 CO2 Chlamydomonas reinhardtii		
ACCESSION	CDNA clone Hc070g06_r 5', mRNA sequence.		
VERSION	AV637259		
KEYWORDS	AV637259.1 GI:10780579		
SOURCE	EST.		
ORGANISM	Chlamydomonas reinhardtii. Chlamydomonas reinhardtii. Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales; Chlamydomonadaceae; Chlamydomonas.		
REFERENCE	1 (bases 1 to 450)		
AUTHORS	Asamizu,E., Mura,K., Kuchio,K., Inoue,Y., Fukuzawa,H., Ohnaya,K., Nakamura,Y. and Tabata,S.		
TITLE	Generation of expressed sequence tags from low-CO2 and high-CO2 adapted cells of Chlamydomonas reinhardtii		
JOURNAL MEDLINE	DNA Res. 7 (5), 305-307 (2000)		
COMMENT	20539644 Contact: Erika Asamizu The First Laboratory for Plant Gene Research Kazusa DNA Research Institute Yama 1532-3, Kisarazu, Chiba 292-0812, Japan Email: asamizue@kazusa.or.jp, URL: http://www.kazusa.or.jp/en/plant/. Location/Qualifiers		
FEATURES	source		
	1..450		
	/organism="Chlamydomonas reinhardtii"		
	/strain="CG"		
	/db_xref="taxon:3055"		
	/clone="HC070g06_r"		
	/note="Vector: pBluescriptII SK-; Site1: EcoRI; Site2: XhoI; The cDNA library was constructed from cells cultured in a medium with bubbling air containing 5% carbon dioxide"		
BASE COUNT	74 a 153 c 148 g 75 t		
ORIGIN			
	Query Match 6.1%; Score 38.6; DB 10; Length 450;		
	Best Local Similarity 50.8%; Pred. No. 9.6;		
	Matches 92; Conservative 0; Mismatches 89; Indels 0; Gaps 0;		
OY	183	CTTGCCATTACTGCTGTGCATGACATTCCAGCCTTCGGTTATGAGGGCGCAAGTGTG	242
Dd	221	CTTGCACCCCCCTGGACGGCTGCTCCATGTGTGACACCACATTCGCCGTGGCACCATCTT	280
OY	243	CGGGATGTACCATGTGCAGGAAGCACTGCTCCAACTCAAGCATTTGTATGAGGACGCGA	302
Dd	281	CGGGCTGTGGCCCGCGCAACAACCTGACCAAATFACACGGCCGCGAGAGGTGGTGGCGG	340
OY	303	CATGATCATGCAACACCCCCGGTGGTGGCTTCCGCTTGGGAGAACAACCTTTCGGCTG	362
Dd	341	CATGGGCACTTAAGGTCCCGCAGCAGTGTTCGATTGCTCCCTGAAGAGAGACCCCGGCTG	400
OY	363	c	363

Db 401 C 401

RESULT 11  
AV637328 451 bp mRNA linear EST 15-DEC-2000  
LOCUS AV637328 Chlamydomonas reinhardtii 5% CO2 Chlamydomonas reinhardtii  
DEFINITION CDNA clone HC071f12\_r 5', mRNA sequence.  
ACCESSION AV637328  
VERSION AV637328.1 GI:10780648  
KEYWORDS EST  
SOURCE Chlamydomonas reinhardtii.  
ORGANISM Chlamydomonas reinhardtii.  
REFERENCE Chlamydomonas reinhardtii.  
AUTHORS Chlamydomonas reinhardtii.  
TITLE Chlamydomonas reinhardtii.  
JOURNAL Chlamydomonas reinhardtii.  
MEDLINE Chlamydomonas reinhardtii.  
COMMENT Chlamydomonas reinhardtii.  
The First Laboratory for Plant Gene Research  
Kazusa DNA Research Institute  
Yana 1532-3, Kisarazu, Chiba 292-0812, Japan  
Email: asamizu@kazusa.or.jp, URL: http://www.kazusa.or.jp/en/plant/  
Location/Qualifiers  
1. 451  
/organism="Chlamydomonas reinhardtii"  
/strain="Cg"  
/db\_xref="taxon:3055"  
/clone="HC071f12\_r"  
/note="Vector: pBluescriptII SK-; Site 1: EcoRI; Site 2: XhoI; The CDNA library was constructed from cells cultured in a medium with bubbling air containing 5% carbon dioxide"

BASE COUNT 74 a 150 c 150 g 77 t

ORIGIN

Query Match 6.1%; Score 38.6; DB 10; Length 451;  
Best Local Similarity 50.8%; Pred. No. 9.6;  
Matches 92; Conservative 0; Mismatches 89; Indels 0; Gaps 0;

QY 183 CTTCGCTTACTGCTCTGCTGACATTCACGCTTCGCTTATGAGTGGCAACGCTG 242  
Db 178 CTTCGACCCCTGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 237  
QY 243 CGGATGTACATGTCACGACGACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 302  
Db 238 CGGCGGTGCGCCGCGGACGACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 297  
QY 303 CATGATCATGACACCCCGGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 362  
Db 298 CATGGCATCTACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 357

QY 363 C 363  
Db 358 C 358

RESULT 12  
AV637643 451 bp mRNA linear EST 15-DEC-2000  
LOCUS AV637643 Chlamydomonas reinhardtii 5% CO2 Chlamydomonas reinhardtii  
DEFINITION CDNA clone HC075g03\_r 5', mRNA sequence.  
ACCESSION AV637643  
VERSION AV637643.1 GI:10780963  
KEYWORDS EST  
SOURCE Chlamydomonas reinhardtii.  
ORGANISM Chlamydomonas reinhardtii.

REFERENCE Chlamydomonas reinhardtii.  
AUTHORS Chlamydomonas reinhardtii.  
TITLE Chlamydomonas reinhardtii.  
JOURNAL Chlamydomonas reinhardtii.  
MEDLINE Chlamydomonas reinhardtii.  
COMMENT Chlamydomonas reinhardtii.  
The First Laboratory for Plant Gene Research  
Kazusa DNA Research Institute  
Yana 1532-3, Kisarazu, Chiba 292-0812, Japan  
Email: asamizu@kazusa.or.jp, URL: http://www.kazusa.or.jp/en/plant/  
Location/Qualifiers  
1. 451  
/organism="Chlamydomonas reinhardtii"  
/strain="Cg"  
/db\_xref="taxon:3055"  
/clone="HC075g03\_r"  
/note="Vector: pBluescriptII SK-; Site 1: EcoRI; Site 2: XhoI; The CDNA library was constructed from cells cultured in a medium with bubbling air containing 5% carbon dioxide"

BASE COUNT 76 a 151 c 149 g 75 t

ORIGIN

Query Match 6.1%; Score 38.6; DB 10; Length 451;  
Best Local Similarity 50.8%; Pred. No. 9.6;  
Matches 92; Conservative 0; Mismatches 89; Indels 0; Gaps 0;

QY 183 CTTCGCTTACTGCTCTGCTGACATTCACGCTTCGCTTATGAGTGGCAACGCTG 242  
Db 196 CTTCGACCCCTGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 255  
QY 243 CGGATGTACATGTCACGACGACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 302  
Db 256 CGGCGGTGCGCCGCGGACGACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 315  
QY 303 CATGATCATGACACCCCGGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 362  
Db 316 CATGGCATCTACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 375

QY 363 C 363  
Db 376 C 376

RESULT 13  
AV634724 453 bp mRNA linear EST 15-DEC-2000  
LOCUS AV634724 Chlamydomonas reinhardtii 5% CO2 Chlamydomonas reinhardtii  
DEFINITION CDNA clone HC037a06\_r 5', mRNA sequence.  
ACCESSION AV634724  
VERSION AV634724.1 GI:10778044  
KEYWORDS EST  
SOURCE Chlamydomonas reinhardtii.  
ORGANISM Chlamydomonas reinhardtii.  
REFERENCE Chlamydomonas reinhardtii.  
AUTHORS Chlamydomonas reinhardtii.  
TITLE Chlamydomonas reinhardtii.  
JOURNAL Chlamydomonas reinhardtii.  
MEDLINE Chlamydomonas reinhardtii.  
COMMENT Chlamydomonas reinhardtii.  
The First Laboratory for Plant Gene Research  
Kazusa DNA Research Institute  
Yana 1532-3, Kisarazu, Chiba 292-0812, Japan  
Email: asamizu@kazusa.or.jp, URL: http://www.kazusa.or.jp/en/plant/.

FEATURES	Location/Qualifiers
SOURCE	1. 453 /organism="Chlamydomonas reinhardtii" /strain="C9" /db_xref="taxon:3055" /clone="HC037406_r" /clone_1lb="Chlamydomonas reinhardtii 5% CO2" /note="Vector: pBluescriptII SK-; Site_1: EcoRI; Site_2: XhoI; The cDNA library was constructed from cells cultured in a medium with bubbling air containing 5% carbon dioxide"
BASE COUNT	76 a 154 c 148 g 75 t
ORIGIN	
Query Match	6.1%; Score 38.6; DB 10; Length 453;
Best Local Similarity	50.8%; Pred. No. 9.6;
Matches 92; Conservative 0; Mismatches 89; Indels 0; Gaps 0;	
OY	183 CTTGGCTTTACGTCGCTGTCTGACCATTCCTCCGCTATGAGGTCGACAGGTGC 242 
DB	203 CTTTCACACCCCTGAGCAGGCTGCTTCATCTGTGCACACCACTTCGCCGGGACCATCTT 262 
OY	243 CGGGATGTAACCACTGTCACGAAACAGACTGCTCCAATCAAGCATTTGTATGACAGCGGA 302 
DB	263 CGGGCTGTGGCCGGCGGACCAAGCTGACCAACATCAACCGGCGCGAGAGGTGCTGCGGG 322 
OY	303 CATGATCATGACGACACCCCGGGGTGGTGGTCCCTTCGCTTGCGGAGAACACTCTTCCCGTCG 362 
DB	323 CATGGGCATCTACGAGGTCCCGGACCGTGTTCGATTTGCCCTGAAGGAGCGCCCGGCTG 382 
OY	363 C 363 
DB	383 C 383
RESULT 14	
AV637050	454 bp mRNA linear EST 15-DEC-2000
LOCUS	AV637050 Chlamydomonas reinhardtii 5% CO2 Chlamydomonas reinhardtii
DEFINITION	CDNA clone HC068804_r 5', mRNA sequence.
ACCESSION	AV637050
VERSION	AV637050.1 GI:10780370
KEYWORDS	EST.
SOURCE	Chlamydomonas reinhardtii.
ORGANISM	Chlamydomonas reinhardtii. Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales; Chlamydomonadaceae; Chlamydomonas.
REFERENCE	1 (bases 1 to 454) Asamizu,E., Miura,K., Kuchino,K., Inoue,Y., Fukuzawa,H., Ohyama,K., Nakamura,Y. and Tabata,S.
AUTHORS	Generation of expressed sequence tags from low-CO2 and high-CO2, adapted cells of Chlamydomonas reinhardtii
TITLE	DNA Res. 7 (3), 305-307 (2000)
JOURNAL	20539644
MEDLINE	Contact: Erika Asamizu The First Laboratory for Plant Gene Research Kazusa DNA Research Institute Yana 1532-3, Kisarazu, Chiba 292-0812, Japan Email: asamizu@kazusa.or.jp, URL: http://www.kazusa.or.jp/en/plant/. Location/Qualifiers
COMMENT	1. 454 /organism="Chlamydomonas reinhardtii" /strain="C9" /db_xref="taxon:3055" /clone="HC068804_r" /clone_1lb="Chlamydomonas reinhardtii 5% CO2" /note="Vector: pBluescriptII SK-; Site_1: EcoRI; Site_2: XhoI; The cDNA library was constructed from cells cultured in a medium with bubbling air containing 5% carbon dioxide"
FEATURES	
SOURCE	
BASE COUNT	82 a 153 c 144 g 75 t
ORIGIN	

Query Match	6.1%	Score 38.6:	DB 10;	Length 456;
Best Local Similarity	50.8%;	Pred. No. 9.6;		
Matches 92;	Conservative 0;	Mismatches 89;	Indels 0;	Gaps 0;
OY	183 CTTGGCTTTCACGTGCTCCTGTGCATCCATTCCACCCTTGCGCTTATAGAGTGGCGCAACGTGTC	242		
Dd	52 CTTCGACCCCCCTGGAGAGGCTGCTCATGTGTGCACACAACTTCGCCGTGGGACCATCTT	111		
OY	243 CGGGATGTACCACTGTACGACAGCACTGCTCCAACTCAAGACTTGTATAGACGACGGGA	302		
Dd	112 CGGCGTGTGGCCCCGGGACAACTGACAACCAATCACCGGCCGCGAGAGGTGGTCCGG	171		
OY	303 CATGATCATGCAACACACCCCAGGTGGTGGCTTCCGCTTGGGAGAACAACACTTCTCCCGTG	362		
Dd	172 CATGGGCACTATCTAGGCTCCCGGACCGTGTTCGATTCGCTTGAAGAGAGCCCGCGCTG	231		
OY	363 C 363			
Dd	232 C 232			
RESULT 15				
LOCUS	AV635382	456 bp	mRNA	linear EST 15-DEC-2000
DEFINITION	AV635382 Chlamydomonas reinhardtii 5% CO <sub>2</sub> Chlamydomonas reinhardtii			
KEYWORDS	CDNA clone HCO45f10_r_5 , mRNA sequence.			
VERSION	AV635382			
KEYWORDS	AV635382.1 GI:10778702			
SOURCE	EST.			
ORGANISM	Chlamydomonas reinhardtii.			
REFERENCE	Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;			
AUTHORS	Chlamydomonadaceae; Chlamydomonas.			
TITLE	1 (bases 1 to 456)			
JOURNAL MEDLINE	Asamizu,E., Mura,K., Kuchio,K., Inoue,Y., Fukuzawa,H., Ohshima,K., Nakamura,Y. and Tabata,S.			
COMMENT	Generation of expressed sequence tags from low-CO <sub>2</sub> and high-CO <sub>2</sub> adapted cells of Chlamydomonas reinhardtii			
FEATURES	DNA Res. 7 (5), 305-307 (2000)			
source	Contact: Erika Asamizu The First Laboratory for Plant Gene Research Kazusa DNA Research Institute Yata 1532-3, Kisarazu, Chiba 292-0812, Japan Email: asamizuekazusa.or.jp, URL:http://www.kazusa.or.jp/en/plant/. Location/Qualifiers			
	1..456			
	/organism="Chlamydomonas reinhardtii"			
	/strain="C9"			
	/db_xref="taxon:3055"			
	/clone="HCO45f10_r"			
	/note="Vector: pBluescriptII SK-; Site 1: EcoRI; Site 2: XhoI; The cDNA library was constructed from cells cultured in a medium with bubbling air containing 5% carbon dioxide"			
BASE COUNT	75 a 158 c 148 g 75 t			
ORIGIN				
* Query Match	6.1%	Score 38.6:	DB 10;	Length 456;
Best Local Similarity	50.8%;	Pred. No. 9.6;		
Matches 92;	Conservative 0;	Mismatches 89;	Indels 0;	Gaps 0;
OY	183 CTTGGCTTTCACGTGCTCCTGTGCATCCATTCCACCCTTGCGCTTATAGAGTGGCGCAACGTGTC	242		
Dd	221 CTTCGACCCCCCTGGAGAGGCTGCTCATGTGTGCACACAACTTCGCCGTGGGACCATCTT	280		
OY	243 CGGGATGTACCACTGTACGACAGCACTGCTCCAACTCAAGACTTGTATAGACGACGGGA	302		
Dd	281 CGGCGTGTGGCCCCGGGACAACTGACAACCACTCACCGGCCGCGAGAGGTGGTCCGG	340		
OY	303 CATGATCATGCAACACACCCCAGGTGGTGGCTTCCGCTTGGGAGAACAACACTTCTCCCGTG	362		

Db 341 CATGGGCACTCTACGGTCCCCGACCGCTTTCGCAATTGCCCTGAGAGACGCCCCGCGTg 400

Qy 363 C 363

Db 401 C 401

Search completed: February 18, 2003, 07:07:42  
Job time : 1698 secs